

‘Lessons learned’

Onderzoek naar de benutting van ‘lessons learned’ evaluaties binnen de Luchtmacht

Open Universiteit Nederland
Faculteit Managementwetenschappen
Public Management

Naam: J. H. van der Veer
Studentnummer: 838670561
Datum: 13 juni 2013

Begeleider/examinator: Mw. Lic. M. Crijns
Medebeoordelaar: Dhr. Drs. G. Arendsen

Samenvatting

De overheid wil een lerende organisatie zijn. Dit 'lerend' vermogen wordt onder andere in de organisatie ingebed door het uitvoeren van evaluaties, die op 'leren' zijn gericht. Als onderdeel van de overheidsorganisatie kent ook de krijgsmacht vele vormen van controle op basis van evaluaties. Naast verantwoording van gevoerd beleid en effectiviteitsmetingen, heeft de krijgsmacht specifieke vormen van evaluatie die inzicht moeten geven in de operationele slagkracht van militaire eenheden. De afgelopen jaren is er – mede ingegeven door internationale ontwikkelingen op dit terrein – steeds meer aandacht voor het 'leren' van deze ervaringen. De krijgsmacht hanteert hierbij de z.g. 'lessons learned' systematiek.

In het huidige, wetenschappelijke denken over evalueren staat 'use' of 'utilization' (benutting) centraal. De doelstelling van het onderzoek is na te gaan of de aard van de resultaten die voortkomen uit het 'lessons learned' proces van invloed is op de wijze van benutting en zo ja, welke vorm van benutting. Daarnaast wordt bezien of benutting ook als zodanig wordt (h)erkend. De centrale vraag die in dit onderzoek wordt beantwoord luidt:

In hoeverre bepaalt de aard van de resultaten die voortkomen uit het 'lessons learned' proces de wijze waarop benutting plaatsvindt?

De aard van de resultaten ('findings') is beschouwd op twee punten. Ten eerste is nagegaan in hoeverre de implementatiecomplexiteit van invloed is op de benutting. Daarnaast is gekeken naar een mogelijk verschil tussen positieve ('best practices') en negatieve resultaten (verbeterpunten). Voor wat betreft de benutting is gekozen voor de drie 'klassieke' vormen van benutting: instrumenteel (impliceert enige vorm van een beslissing/besluit), conceptueel (genereert ideeën, begrip en daarmee discussie) en symbolisch (of politiek) (leidt tot ondersteuning van argumenten of rechtvaardiging van (geen) actie).

De inzet van luchtmachteenheden kan kortstondig zijn of zich uitstrekken over een langere periode, zoals dat bijvoorbeeld het geval is in Afghanistan. Het personeel dat behoort tot deze eenheden wordt als regel uitgezonden voor een periode van vier tot zes maanden. Dit betekent dat voor langlopende inzet – zoals de inzet in Afghanistan – meerdere detachementen dezelfde taak uitvoeren. Het onderzoek bestudeert drie opeenvolgende evaluaties van F-16 detachementen in Afghanistan gedurende de periode van eind 2010 tot eind 2011.

Voor dit onderzoek is als onderzoeksstrategie gekozen voor de gevalsstudie. Gevalsstudies dienen aan drie voorwaarden te voldoen. Ten eerste is de methode vooral geschikt voor 'hoe' en 'waarom' vragen (Yin, 2009). De centrale vraag van dit onderzoek voldoet aan die voorwaarde. De tweede voorwaarde is de mate van controle over de uitkomst van het resultaat en/of gedragingen. Invloed op de evaluatieresultaten was niet mogelijk. Dit betekent dat het onderzoek ook aan deze tweede voorwaarde voldoet. De derde voorwaarde betreft de focus op huidige versus historische gebeurtenissen. Gevalsstudie beschouwt de huidige situatie. Het onderzoek richt zich op een hedendaags verschijnsel en voldoet daarom ook aan de derde voorwaarde.

Er zijn drie onderzoekbronnen gebruikt. Ten eerste de 'lessons learned' database. Deze database bevat de resultaten van alle evaluaties die hebben plaatsgevonden in het kader van het 'lessons

learned' proces. Een tweede bron van informatie werd gevormd door de evaluatieverslagen zelf. Deze verslagen zijn gebruikt om de evaluatieresultaten dieper te analyseren en in hun context te plaatsen. De derde bron van informatie waren de gevoerde gesprekken. Deze gesprekken hadden een tweeledig doel. Ten eerste als additionele informatiebron en daarnaast als toetsingsmechanisme voor de geanalyseerde data.

Er is getracht om een zo valide en betrouwbaar mogelijk onderzoek uit te voeren. Toch kent het onderzoek een aantal beperkingen. Hoewel de onderzochte cases met zorg zijn gekozen en deze keuze representatief lijkt, is dit niet met zekerheid vast te stellen. Daarvoor zouden alle resultaten uit de database moeten worden onderzocht. Ten tweede heeft een beoordeling en indeling van meer of minder complexe resultaten plaatsgevonden. Ook de toekenning van het soort benutting, is uitgevoerd door eigen analyse en aan de hand van de gevoerde gesprekken. Deze vorm van indeling en oordeelsvorming heeft per definitie een subjectief karakter. Ten derde is de aard van de resultaten slechts op twee dimensies beschouwd, wat mogelijk een vertekend beeld oplevert. Tot slot worden in het onderzoek uitspraken gedaan op grond van slechts een beperkt aantal evaluatieresultaten. Generalisatie van de uitkomsten van het onderzoek dient dan ook met de nodige voorzichtigheid en bescheidenheid plaats te vinden.

Het onderzoek heeft de volgende resultaten opgeleverd. Ten aanzien van de implementatiecomplexiteit bleek dat minder complexe resultaten uitsluitend hebben geleid tot instrumentele benutting. Voor meer complexe resultaten was er in de meeste gevallen weliswaar sprake van instrumentele benutting, maar in een aantal gevallen kon ook conceptuele benutting worden aangetoond en bij één resultaat was er sprake van symbolische benutting. Daarnaast zijn er resultaten gevonden, waarbij er per resultaat meerdere vormen van benutting kon worden vastgesteld.

Het onderzoek heeft verder aangetoond dat alle positieve resultaten uit de cases instrumenteel zijn benut. Bij de negatieve resultaten zijn ook andere vormen van benutting gevonden. Er was bij een aantal resultaten sprake van conceptuele benutting en in één geval van symbolische benutting. Daarnaast kon worden vastgesteld dat positieve resultaten vaker leiden tot benutting dan negatieve resultaten.

De aandacht voor positieve resultaten lijkt dus terecht. Het onderzoek lijkt de kracht van het innovatieve vermogen op de 'werkvloer' te bevestigen. Hierbij betreft het niet alleen het verbetervoorstel op zich. In veel gevallen heeft ook de implementatie van het voorstel al plaatsgevonden. Hierdoor resteert alleen een goede borging.

Als 'bijvangst' van het onderzoek is geconstateerd dat er een groot verschil bestaat tussen de mate waarin resultaten leiden tot 'geleerde lessen' en de mate waarin benutting van resultaten kan worden vastgesteld. Het 'lessons learned' proces leidt slechts voor een beperkt deel van de resultaten tot gebruik ('geleerde les'). Daarnaast betreffen deze 'lessen' vaak kleine, eenvoudig te implementeren resultaten. Wordt echter gekeken naar de benutting doemt een heel ander beeld op. Bij een groot deel van de resultaten kan benutting worden aangetoond en blijken ook meer complexe resultaten te leiden tot enige vorm van benutting. (H)erkenning van deze benutting lijkt echter slechts beperkt plaats te vinden.

Op grond van de resultaten van dit onderzoek zou dan ook een ander beeld ten aanzien van de relevantie van dit type evaluaties kunnen ontstaan. Eerst binnen de 'lessons learned' organisatie zelf vanuit het besef dat er ook zonder 'geleerde les' sprake kan zijn van benutting. Daarnaast zou – gezien de kracht van positieve resultaten – de focus op 'best practices' verder kunnen worden geëntameerd.

Hoewel er binnen overheidsorganisaties (en dus ook de Luchtmacht) aandacht is voor het verbeteren van de eigen prestatie, wordt echter ook geconstateerd dat – als gevolg van de recente bezuinigingen – er een tendens waarneembaar is, de capaciteit om deze programma's uit te voeren te beperken. Goede, regelmatige communicatie met de diverse 'stakeholders' lijkt essentieel voor het welslagen van het proces per resultaat én de waardering voor het programma als geheel.

Inhoudsopgave

Samenvatting	1
Inhoudsopgave	4
1 Inleiding	5
1.1 Introductie	5
1.2 Verwachtingen	5
1.3 Casus	6
1.4 Wetenschappelijke relevantie	6
1.5 Maatschappelijke relevantie	6
1.6 Probleemstelling	7
1.7 Redenen voor het inperken van het onderwerp	8
1.8 Leeswijzer – structuur van de scriptie	8
2 Theoretisch kader	9
2.1 Inleiding	9
2.2 Wat wordt benut?	9
2.3 Hoe wordt benut?	10
2.4 Factoren die benutting beïnvloeden	11
2.5 Factoren in de categorie ‘Evaluation implementation’	13
2.6 Findings – de aard van resultaten	14
2.7 Typen benutting	15
2.8 Samenvatting	17
3 Methodologie	18
3.1 Inleiding	18
3.2 De gevalsstudie	18
3.3 Ontwerp van een gevalsstudie	18
3.4 Datacollectie	20
3.5 Kwaliteitseisen	20
3.6 Keuzes voor het onderzoek verklaard	21
3.7 Samenvatting	23
4 Resultaten van het onderzoek	25
4.1 Inleiding	25
4.2 In hoeverre worden de resultaten die voortkomen uit een evaluatie ‘geleerde lessen’?	25
4.3 In hoeverre is de ‘aard van de resultaten’ van invloed op de mate van de voortgang in het proces?	28
4.4 Op welke wijze zijn resultaten die voortkomen uit het ‘lessons learned’ proces benut?	29
4.5 Zijn er resultaten waarvan benutting niet als zodanig is (h)erkend?	31
4.6 Conclusie	33
5 Conclusies, discussie en aanbevelingen	35
5.1 Inleiding	35
5.2 Conclusies	35
5.3 Verwachtingen	36
5.4 Relevantie	36
5.5 De onderzoeksresultaten ten opzichte van het theoretisch kader	36
5.6 Betrouwbaarheid en validiteit van onderzoek	37
5.7 Beperkingen van het onderzoek	38
5.8 Discussie naar aanleiding van het onderzoek	38
5.9 Aanbevelingen naar aanleiding van het onderzoek	38
Literatuurlijst	40
Bijlage A: Explicatie ‘lessons learned’ systematiek	43
Bijlage B: Tabellen onderzochte resultaten	45
Bijlage C: Gesprekverslag Hr. A.P. van Hees	48
Bijlage D: Gesprekverslag Hr. E.F. op den Kamp	52
Bijlage E: Gesprekverslag Hr. H. van der Leest	54

1 Inleiding

1.1 **Introductie**

1.1.1 *Goed openbaar bestuur*

De overheid is verantwoordelijk voor goed openbaar bestuur. In het rapport van de Rekenkamer¹ wordt inhoud gegeven aan de definitie 'goed openbaar bestuur'. Als controlerend orgaan richt de rekenkamer zich vooral op twee kenmerken: het functioneren en het presteren van de overheid. Ten aanzien van het functioneren van de overheid wordt een onderscheid gemaakt tussen transparantie en publieke verantwoording. Voor wat betreft de prestatie moet het beleid doeltreffend, doelmatig zijn en dient het aan te sluiten op de uitvoering. Het uitvoeren van beleidsevaluaties kan bijdragen aan het verbeteren van de prestatie. Daarnaast dient ook een intern proces plaats te vinden. Het gaat om het verbeteren van de eigen prestatie door te 'leren'. De kwaliteitszorgsystemen, de verantwoording en het toezicht moeten op 'leren' zijn gericht. De evaluatiefunctie is ingebed in de organisatie en biedt daardoor een waarborg voor een lerende organisatie.

1.1.2 *Evaluatie en de krijgsmacht*

Als onderdeel van de overheidsorganisatie kent ook de krijgsmacht vele vormen van controle op basis van evaluaties. Naast verantwoording voor gevoerd beleid en effectiviteitsmetingen, kent de krijgsmacht specifieke vormen van evaluatie die inzicht moeten geven in de operationele slagkracht van militaire eenheden. De afgelopen jaren – mede ingegeven door internationale ontwikkelingen op dit terrein – is er steeds meer aandacht voor het 'leren' van de ervaringen bij operationele inzet van de krijgsmacht. De door de krijgsmacht gehanteerde methodiek is gebaseerd op de methodiek die door de NAVO is beschreven in het 'The NATO Lessons Learned Handbook'².

Binnen de Luchtmacht wordt sinds drie jaar invulling gegeven aan de opdracht van de Commandant der Strijdkrachten (CDS) om door middel van evaluaties te leren van de inzet van Defensie. Deze opdracht is verwoord in de z.g. CDS-aanwijzing A-1202. In de aanwijzing valt te lezen: *'Tijdens de operationele inzet van de krijgsmacht worden waardevolle lessen opgedaan'. ... 'Defensie wil een lerende organisatie zijn. Daarom dienen zowel positieve ('best practices') als negatieve opgedane ervaringen na validatie en analyse, zoveel mogelijk te worden ontwikkeld naar concrete maatregelen die leiden tot een effectievere en efficiëntere inzet van de krijgsmacht. Dit proces wordt aangemerkt als het 'Lessons Learned' proces Operaties'*³.

Uit bovenstaande tekst blijkt een focus op twee elementen. Ten eerste wordt gesproken over 'concrete maatregelen', hetgeen benutting impliceert. Ten tweede wordt met nadruk gesteld dat het niet alleen gaat om negatieve verbeterpunten, maar ook om positieve ervaringen ('best practices').

1.2 **Verwachtingen**

Naar verwachting zal de benutting van betrekkelijk eenvoudig te implementeren aanbevelingen voornamelijk instrumenteel van aard zijn. Ten aanzien van complexere aanbevelingen zal er wellicht ook sprake zijn van andere vormen van benutting.

¹ Essentia van goed openbaar bestuur, uitgave Rekenkamer

² NATO lessons Learned Handbook', uitgave NATO's JOINT ANALYSIS AND LESSONS LEARNED CENTRE, september 2011

³ Uit: CDS Aanwijzing A-1202 1 juli 2010.

Dit vermoeden wordt versterkt door de visie van Brober en Bartlett (2004: 367). In hun artikel verwijzen ze naar een onderzoek dat is uitgevoerd door Preskill in 1984. Deze studie wees uit dat evaluatieresultaten zowel instrumenteel als conceptueel werden gebruikt. Een veel groter percentage van de benutting van de resultaten was echter instrumenteel van aard. Het bleek lastig benutting van conceptuele aard te duiden en als zodanig te herkennen.

Het tweede onderscheid betreft de aard van de resultaten. Er dient aandacht te zijn voor zowel positieve als negatieve resultaten. Weiss e.a. stellen dat deze vorm van evaluatie zelfs uitnodigt om druk uit te oefenen om de 'best practices identified' breed toe te passen (Weiss e.a 2005. p. 26). Of dit invloed heeft op de benutting, valt vooralsnog niet te voorspellen.

1.3 Casus

Nederlandse luchtmachteenheden zijn de afgelopen twintig jaar regelmatig ingezet als bijdrage van Nederland aan VN- of NAVO vredesoperaties. Zo'n inzet kan kortstondig zijn (bijvoorbeeld de inzet van F-16 gevechtsvliegtuigen tijdens de NAVO-inzet in Libië) of zich uitstrekken over een langere periode, zoals dat bijvoorbeeld het geval is in Afghanistan. Het personeel dat behoort tot deze eenheden wordt als regel uitgezonden voor een periode van vier tot zes maanden. Dit betekent dat voor langlopende inzet – zoals de inzet in Afghanistan – meerdere detachementen dezelfde taak uitvoeren. Ieder detachement moet na afloop van zijn missie een evaluatieverslag schrijven. Sinds 2010 moet het verslag conform de 'lessons learned' systematiek zijn ingericht. Doel is steeds van de ervaringen te leren en zo nodig deze ervaringen te vertalen naar verbeteringen in werkwijze, organisatie en/of optreden. De onderzochte cases betreffen drie opeenvolgende evaluaties van F-16 detachementen (ATF 15, 16 en 17) in Afghanistan gedurende de periode van eind 2010 tot eind 2011.

1.4 Wetenschappelijke relevantie

Over benutting van evaluaties is zeer veel geschreven. C.A. Christie (2007:8) schrijft bijvoorbeeld *'Evaluation utilization is arguably the most researched area of evaluation'*. Echter, in hetzelfde artikel valt ook te lezen dat veel hypothesen ten aanzien van benutting van evaluaties stoelen op een smalle empirische basis. Dit onderzoek richt zich op de wijze van benutting en toetst de daadwerkelijke resultaten in termen van nut voor de organisatie. Hiermee wordt een bijdrage geleverd aan het verbreden van genoemde empirische basis.

1.5 Maatschappelijke relevantie

Binnen overheidsorganisaties is een toenemende belangstelling voor de effecten van de eigen activiteiten waarneembaar. Deze interesse komt niet alleen voort uit de plicht om (politieke) verantwoording over deze activiteiten af te leggen, maar is ook intern gericht om de effectiviteit van het eigen beleid (en de organisatie) te meten. Daarnaast is er de wens om de eigen organisatie te verbeteren. Evaluatie als middel om tekortkomingen te signaleren en hiervan te leren wordt steeds vaker toegepast. Toch kan ook worden geconstateerd dat bij overheidsorganisaties, als gevolg van de recente bezuinigingen, de capaciteit om deze programma's uit te voeren wordt beperkt. De bevindingen uit dit onderzoek kunnen een bijdrage leveren aan de discussie over de relevantie van dit type evaluaties voor overheidsorganisaties.

1.6 Probleemstelling

1.6.1 Doelstelling

De doelstelling van het onderzoek is na te gaan of de aard van de resultaten, die voortkomen uit het 'lessons learned' proces, van invloed is op de wijze van benutting. In het verlengde hiervan ligt de vraag in hoeverre er – naast instrumentele – ook sprake is van andere vormen van benutting. En zo ja, of deze vormen van benutting ook als zodanig worden (h)erkend.

1.6.2 Centrale vraag

In hoeverre bepaalt de aard van de resultaten, die voortkomen uit het 'lessons learned' proces, de wijze waarop benutting plaatsvindt?

1.6.3 Deelvragen

Om de centrale vraag te beantwoorden zijn een aantal deelvragen geformuleerd, die niet alleen een theoretisch fundament onder het onderzoek leggen maar ook het onderzoek zelf richting geven.

Als algemene inleiding in het onderwerp wordt in het volgende hoofdstuk een theoretisch kader geschapen als achtergrondinformatie en ter ondersteuning van het onderzoek. Er wordt ten eerste ingegaan op de vraag: **Wat wordt verstaan onder benutting van evaluaties?** Ook de tweede deelvraag maakt onderdeel uit van het theoretisch kader. Het antwoord op de vraag beschrijft de factoren die een rol spelen bij evaluatiebenutting. Hierbij wordt aan de hand van het overzicht van Johnson (Johnson e.a. 2009) – met als basis het Cousins en Leithwood raamwerk – invulling gegeven aan de vraag: **Welke factoren spelen een rol bij de benutting van evaluaties?** Met de beantwoording van de vraag: **Wat wordt bedoeld met de 'aard van de resultaten'?** wordt nader ingegaan op de factor 'findings', zoals genoemd in het raamwerk van Cousins en Leithwood. Deze factor wordt eerst uitgewerkt in termen van het relatieve belang ten opzichte van andere factoren. Daarna wordt uitgelegd wat wordt bedoeld met de 'aard van de resultaten'. De vierde deelvraag luidt: **Welke vormen van benutting zijn er?** Het antwoord op deze vraag richt zich op de wijze van benutting en vormt de laatste bouwsteen van het theoretisch kader. Daarnaast passeren de diverse typen (of vormen) van evaluatiebenutting de revue en wordt de keuze voor de – in het onderzoek – gehanteerde vormen van benutting verklaard.

De volgende vragen geven richting aan het onderzoek, waarvan de resultaten zijn ondergebracht in hoofdstuk vier. De eerste vraag biedt een overzicht van de bereikte evaluatieresultaten. Voor dit overzicht is gebruik gemaakt van de 'lessons learned' database en de gesprekken met de betrokken 'lessons learned' functionaris (zie bijlage c). De beantwoording is hoofdzakelijk kwantitatief van aard. Hoofdzakelijk, omdat er bij de analyse toch sprake is van een (subjectieve) beoordeling en toetsing. De vraag luidt als volgt: **In hoeverre worden de resultaten, die voortkomen uit een evaluatie, 'geleerde lessen'?** De tweede vraag: **In hoeverre is de 'aard van de resultaten' van invloed op de mate van voortgang in het proces?** Met deze vraag wordt nagegaan of er een onderscheid is te duiden tussen de 'aard van de resultaten' en de mate waarin dit leidt tot 'lessons learned' c.q. 'best practices'. Ook deze vraag is beantwoord door gebruik te maken van de 'lessons learned' database en de gesprekken met de 'lessons learned' functionaris (zie bijlage c). De derde vraag gaat na in hoeverre de 'geleerde lessen' daadwerkelijk worden benut en zo ja, op welke wijze benutting plaatsvindt. Voorwaarde om deze vraag deugdelijk te beantwoorden is – naast de gegevens uit de database – de context van het resultaat. Hiervoor zijn de achterliggende verslagen en rapporten gebruikt. Bovendien zijn tijdens de gesprekken gerichte vragen gesteld om eenduidigheid in de

beantwoording zeker te stellen. De vraag luidt: **Op welke wijze zijn resultaten, die voortkomen uit het 'lessons learned' proces, benut?**

In het verlengde van de voorgaande vraag ligt de onderzoeksvraag: **Zijn er resultaten waarvan benutting niet als zodanig is (h)erkend?** Bij de beantwoording van deze vraag zijn de uitkomsten van de verschillende gesprekken leidend.

Vervolgens wordt in hoofdstuk vijf gezien in hoeverre de bevindingen uit het onderzoek aansluiten bij verklaringen, zoals die voorkomen in de theorie: **In hoeverre sluiten de bevindingen van het onderzoek aan bij de resultaten uit vergelijkbaar onderzoek en/of de theoretische onderbouwing?**

Gezien de aard van het onderzoek wordt er ook een uitspraak gedaan over de kwaliteit van het onderzoek. Naast de uiteenzetting in hoofdstuk drie met betrekking tot de gevolgde methodologie, wordt in de conclusie nogmaals een oordeel gegeven over de kwaliteit van het onderzoek. De hierbij gehanteerde vraag is: **Op welke wijze is het resultaat van het onderzoek methodologisch verantwoord?**

1.7 Redenen voor het inperken van het onderwerp

1.7.1 De factor 'findings'

Het raamwerk van Cousins en Leithwood gaat uit van twee categorieën van factoren. De eerste categorie betreft de karakteristieken van evaluatie-implementatie. De tweede categorie omvat de karakteristieken van besluitvorming- of beleidskader. De 'lessons learned' systematiek is gericht op het implementeren van resultaten uit evaluaties met als doel het functioneren van de organisatie te verbeteren. Er is daarom gekozen om de factor 'findings' uit de categorie 'Evaluation implementation' voor dit onderzoek te gebruiken. Daarnaast is – mede met het oog op de doelstelling van het onderzoek – nagegaan of de aard van die resultaten invloed heeft op de wijze waarop deze resultaten worden benut.

1.7.2 Aard van de resultaten

Ten aanzien van de aard van de resultaten is ten eerste gekeken naar een mogelijk verschil tussen positieve ('best practices') en negatieve resultaten (verbeterpunten). Daarnaast is de waarde van de bevindingen voor de besluitvorming in termen van implementatiecomplexiteit onderzocht.

1.7.3 Benutting

Voor wat betreft de wijze van benutting richt het onderzoek zich op de drie 'klassieke' vormen van benutting. De instrumentele, conceptuele en symbolische (of politieke) benutting van de resultaten. Andere vormen van benutting liggen minder voor de hand als gekeken wordt naar het type van evaluatie waartoe de 'lessons learned' systematiek behoort. Daarbij zijn andere vormen van benutting in veel gevallen slechts variaties op de genoemde 'klassieke' drie.

1.8 Leeswijzer – structuur van de scriptie

Om bovengenoemde vragen te beantwoorden, is in het volgende hoofdstuk – op grond van een selectie uit de beschikbare wetenschappelijke literatuur – een relevant kader geschetst. Hoofdstuk drie is gewijd aan de gehanteerde methodologie. De resultaten van het onderzoek zijn verzameld in hoofdstuk vier. Het geheel krijgt zijn afronding in hoofdstuk vijf met conclusies en aanbevelingen.

2 Theoretisch kader

2.1 Inleiding

In het huidige denken over evalueren staat 'use' of 'utilization' centraal. *'Evaluation use or evaluation utilization refers to the way in which an evaluation and information from the evaluation impacts the program that is being evaluated'* (Alkin en Taut, 2003. p.1). Baanbrekend was het werk van Daniel Stufflebeam (in eerste instantie met Egon Guba) en het werk van Joseph Wholey, die zich oorspronkelijk richtte op evaluatie en besluitvorming. Het werk van de theoretici richt zich voornamelijk op de manier waarop de evaluatie van informatie zal worden gebruikt en op degenen die de informatie gebruiken (Alkin en Christie 2004). In de afgelopen jaren hebben wetenschappers in toenemende mate gevraagd om onderzoek naar de invloed en benutting van evaluaties (Henry en Mark, 2003b, Scriven, 2007). Henry en Mark voegen daaraan toe dat *'there is a serious shortage of rigorous, systematic evidence that can guide evaluations or that evaluators can use for self-reflection or for improving their next evaluation'* (Mark en Henry 2003b, p.69).

Dit hoofdstuk schetst een aantal voor het onderzoek relevante bevindingen uit de wetenschappelijke literatuur. Het beoogt een kader te scheppen op grond waarvan het onderzoek kan plaatsvinden.

2.2 Wat wordt benut?

Een eerste vraag die opdoemt is de vraag: 'Wat is het dat wordt benut?' Weiss (1998) beantwoordt die vraag en noemt vijf mogelijkheden:

- a. De bevindingen en aanbevelingen. Door evaluatie geproduceerde bevindingen over een programma, de processen en resultaten van het programma, waarvan de beoordelaars verwachten dat de gebruikers de bevindingen toepassen.
- b. Ideeën en generalisaties uit de evaluatie, ook als deze niet direct voor het specifieke geëvalueerde object van toepassing zijn.
- c. Het gegeven dat evaluatie gebruikt kan worden als rechtvaardiging (of zelfs in-actie). De evaluatie geeft een aura van legitimiteit.
- d. De focus van het onderzoek, in het bijzonder de gebruikte maatregelen (teaching to the test).
- e. De opzet van het onderzoek.

Evaluatiedata of -resultaten kunnen voor verschillende doelen worden gebruikt. Echter, de meest gebruikte indeling in de literatuur is de verbetering van trainingsprogramma's (formatieve evaluatie) en evaluaties waarbij het doel controle, verantwoording en/of voortzetting van programma's (summatieve evaluatie) is (Patton, 1997; Russ-Eft en Preskill, 2001; Worthen, Sanders en Fitzpatrick, 1997).

Bovenstaande mogelijkheden richten zich primair op benutting van evaluatieresultaten. Het antwoord op de vraag lijkt dan ook niet volledig. In 1997 introduceert Patton de term 'process use' en definieert deze term als volgt: *'individual changes in thinking and behavior and program or organizational changes in procedures and culture that occur among those involved in evaluation as result of the learning that occurs during the evaluation process'*. In 1998 gebruikt hij de volgende omschrijving: *'... defined process use as relating to and being indicated by individual changes in thinking and behavior that occur among those involved in evaluation as a result of the learning that*

occurs during the evaluation process.' (Patton, 1998: 225). De term 'proces use' wordt gebruikt wanneer zich een gedeeld begrip van de problemen op het gebied van de onderzochte maatregelen ontwikkelt onder degenen die betrokken zijn bij het evaluatieproces en/of er een netwerk van relaties met daarbij behorende discussies ontstaat.

2.3 Hoe wordt benut?

2.3.1 Verandervermogen

Over het vermogen van organisaties om te 'leren' en als gevolg daarvan te 'veranderen' zijn bibliotheken vol geschreven. Toch is het van belang voor onderzoek naar benutting van evaluatieresultaten, om kort stil te staan bij het vermogen van organisaties om verbeteringen in de organisatie te verwerken, te implementeren of positieve resultaten organisatiebreed toe te passen. In een artikel van Werkman, Boonstra en Bennebroek Gravenhorst (2001), wordt een overzicht geboden en aandacht gegeven aan het belang van terugkoppeling van onderzoeksresultaten aan organisaties in verandering. Zij stellen in de inleiding dat het oorspronkelijke doel van de verandering, het oplossen van een probleem of het introduceren van een verbetering, dikwijls niet wordt gerealiseerd. Veel veranderingsprogramma's slagen er niet in de verandering succesvol af te ronden. Om te bezien in hoeverre een organisatie in staat is te veranderen noemen zij zes kenmerken:

- a. **Doelen en strategie.** Een duidelijke missie en een heldere strategische koers zijn essentieel voor het schetsen van een toekomstbeeld in een organisatie en om de leden van de organisatie een toekomstperspectief te bieden. Het is van belang dat de doelen bekend zijn in een organisatie en dat mensen het eens zijn over deze doelen. Een heldere bedrijfsstrategie is van belang om verwarring en misverstanden te voorkomen over de doelen en de richting van de veranderingen.
- b. **Technologie.** Dit kenmerk betreft de beschikbaarheid van technische hulpmiddelen, de duidelijkheid over het gebruik van die hulpmiddelen en de informatievoorziening van informatiesystemen over de werkkuitvoering. De bijdrage van de technologie varieert van het verbeteren van de flexibiliteit en het innovatievermogen van de organisatie tot controle op bedrijfsprocessen en werkkuitvoering. Als zodanig is technologie te beschouwen als een relevante conditie die veranderingsmogelijkheden creëert en begrenst.
- c. **Structuur.** Structuur verwijst naar de organisatie van het werk en de relaties tussen verschillende eenheden in een organisatie. Een onderscheid kan worden gemaakt tussen mechanische en organische structuren. Mechanische structuren zijn formeel en star, en veranderingsprocessen worden door deze starheid bemoeilijkt. Organische structuren, met een flexibele en innovatieve organisatiestructuur, kunnen organisatieveranderingsprocessen echter vergemakkelijken.
- d. **Cultuur.** De cultuur van een organisatie beschouwt de innovativiteit en ontwikkelgerichtheid van de organisatie, de mensgerichtheid van managers en de mogelijkheden en voorwaarden voor het samenwerken in teams of afdelingen. Cultuurverschillen tussen afdelingen kunnen leiden tot een sterke gerichtheid op de eigen afdeling die het gezamenlijk analyseren van de problemen en werken aan verandering kan bemoeilijken.
- e. **Het werk.** Deze kenmerken omvatten de mate waarin mensen tevreden zijn met hun werk en de mate waarin de organisatie mogelijkheden voor hen creëert om hun kennis en vermogens in hun werk toe te passen en zichzelf te ontwikkelen. Een open omgeving, gericht op leren en ontwikkelen van mensen binnen de organisatie stimuleert hun denkvermogen en slagvaardigheid. Dit kan de slagvaardigheid in veranderingsprocessen vergroten.

- f. **Politieke relaties.** Politieke relaties verwijzen naar de balans tussen autonomie en wederzijdse afhankelijkheid van groepen en individuen in organisaties. Het beschrijft hoe de centraliteit van groepen en afdelingen macht stimuleert, maar uiteindelijk kan leiden tot sterke competitie en non-coöperatief gedrag. Sluimerende of verborgen machtsconflicten tussen onderdelen in de organisatie kunnen tijdens het veranderingsproces leiden tot verstarring van de besluitvorming en onwilligheid om samen te werken.

2.3.2 *Evaluatieresultaten en benutting*

Hoewel bovenstaande opsomming is geformuleerd met ingrijpende organisatieveranderingen of trajecten voor ogen, wordt duidelijk dat implementatie van veranderingen in organisaties vele uitdagingen met zich mee brengt. Dat deze uitdagingen zich niet alleen beperken tot genoemde grootschalige veranderingen, maar ook tot de benutting van evaluatieresultaten, wordt duidelijk door het belang dat wordt toegekend aan 'stakeholders' (belanghebbenden) binnen de theorie van evaluatiebenutting. Algemeen wordt verondersteld dat de benutting van evaluaties gediend is bij betrokkenheid van 'stakeholders' (Christie, 2003; Greene, 1988; Turnbull, 1999). Deze betrokkenheid wordt van een hogere kwaliteit, als de beoordelaar de betrokken 'stakeholders' vroeg identificeert en bij de evaluatie blijft betrekken (Reineke 1991). Fleischer en Christie (2009) stellen dat de betrokkenheid van 'stakeholders' vooral een belangrijk aspect is bij evaluatievormen waarin beoordelaars participeren en waar wordt samengewerkt. Dergelijke rollen worden eveneens genoemd bij evaluaties die een lerend doel hebben (Rossman and Rallis 2000 p. 67).

Ook Shula en Cousins (1997) noemen dit verband (p.198): *'A growing familiarity with the pivotal nature of program contexts was also evident in a third stream of inquiry that focused on the relationship between use and organizational structures and processes (Jenlink, 1994; Mathison, 1994; Owen & Lambert, 1995; Preskill, 1991, 1994a; Preskill & Torres, 1996; Torres, Preskill & Piotnek, 1996). The complexity of bureaucratic levels within the organization, the lines of communication within and across these levels, and the dominant mechanisms for framing the meaning of evaluation information all contributed to the potential utility of evaluation findings'.*

2.4 **Factoren die benutting beïnvloeden**

In deze paragraaf wordt een overzicht gegeven van de diverse factoren die invloed hebben op de benutting van evaluaties. Als basis voor het overzicht is het raamwerk van 'Cousins en Leithwood' (Cousins en Leithwood, 1986) gebruikt. Zij gaan uit van twee categorieën, te weten: karakteristieken van evaluatie-implementatie en karakteristieken van besluitvorming of beleidskader.

Dit raamwerk is aangevuld met het werk van Johnson (Johnson e.a. 2009). Johnson heeft onderzoek gedaan naar evaluatiebenutting in de periode van 1986 tot 2005. Aan de hand van 41 empirische studies (die voldeden aan een aantal minimale kwaliteitseisen) is een overzicht tot stand gekomen waarin deze studies zijn gecategoriseerd. Hierbij heeft het raamwerk van Cousins en Leithwood als basis gediend. Naast de categorieën die worden gebruikt in het raamwerk wordt door Johnson nog een nieuwe categorie geïntroduceerd, te weten 'stakeholder involvement'. De resultaten van het onderzoek van Johnson binnen deze categorie zijn overigens eenduidig, de betrokkenheid van 'stakeholders' draagt voor alle factoren positief bij aan het nut van de evaluatie (waarmee Johnson aansluit bij de eerder genoemde bevindingen in paragraaf 2.3.2.).

Een tweede aanvulling van Johnson betreft het toevoegen van de factor 'evaluator competence' in de categorie 'evaluation implementation'. De bekwaamheid van de beoordelaar is door Johnson als

aparte factor toegevoegd, omdat in diverse onderzoeken (6 van de 41) karakteristieken van de beoordelaar worden genoemd als een belangrijke invloed op de benutting van de evaluatie. Hoewel dit voor een deel al ligt besloten in de factor 'credibility' wordt onder 'evaluator competence' meer gekeken naar het vermogen van de beoordelaar om de benutting van de evaluatie positief te beïnvloeden.

Naast het raamwerk en de aanvullingen door Johnson wordt de derde bouwsteen van het overzicht gevormd door Patton's 'process use'. 'Process use' is opgenomen in dit overzicht omdat het niet moet worden gezien als een nieuwe vorm van benutting. Het legt de nadruk op het effect van deelname aan het evaluatieproces. Patton zegt hierover: *'ways in which being engaged in the processes of evaluation can be useful quite apart from the findings that may emerge from those processes'* (Patton, 1997 p.88). Voor de bepaling van de factoren in deze categorie is gekozen voor de indeling zoals die is gemaakt door Foss (Foss e.a. 2002) te weten: leren te leren, het opbouwen van een professioneel netwerk, het creëren van gedeelde opvattingen, het verstevigen van het project en het verhogen van de moraal.

De genoemde bouwstenen zijn ondergebracht in onderstaand overzicht.

Evaluatie implementatie	besluitvorming- of beleidskader	Stakeholder involvement	Process use
Evaluation quality	Information needs	Involvement with commitment or receptiveness to evaluation	leren te leren
Credibility	Decision characteristics	Involvement with communication quality	opbouwen professioneel netwerk
Relevance	Political climate	Direct stakeholder involvement	creëren gedeelde opvattingen
Communication quality	Competing information	Involvement with credibility	verstevigen project en verhogen moraal
Findings	Personal characteristics	Involvement with findings	
timelessness	Commitment and/or receptiveness to evaluation	Involvement with relevance	
		Involvement with personal characteristics	
		Involvement with decision characteristics	
		Involvement with information needs	
Evaluator competence			

2.5 Factoren in de categorie ‘Evaluation implementation’

Gezien de onderzoeksvraag is er voor gekozen de factoren, zoals die voorkomen in de kolom ‘evaluation implementation’, nader toe te lichten en uit te werken.

2.5.1 *Communication quality*

De kwaliteit van de communicatie rondom evaluaties heeft te maken met de helderheid en de frequentie van de rapportages. Hierbij speelt de mate waarop de beoordelaar resultaten kan verdedigen en uitdragen een rol. Daarnaast omvat deze factor het type van aanbevelingen en de manier waarop de communicatie tussen de geëvalueerde en de beoordelaar is georganiseerd. In het overzicht van Johnson wordt deze factor elf keer genoemd, vaak als één van de belangrijkste factoren in relatie tot benutting. Aanbevelingen die gedetailleerd zijn, actie mogelijk maken en gestoeld zijn op bewijzen, dragen positief bij aan de benutting van de evaluatie. Opmerkelijk is dat er in twee gevallen geen relatie is gevonden.

2.5.2 *Timeliness*

Deze factor gaat in op de snelheid waarmee het rapport na de evaluatie gereed is en wordt aangeboden. In zes gevallen is een positieve relatie gevonden tussen timing en benutting. Deze relatie werd in één studie niet belangrijk gevonden.

2.5.3 *Evaluation quality*

De kwaliteit van de evaluatie wordt hier beschouwd in termen van proces, gebruikte methoden, diepgang en het type evaluatiemodel. Verschillende studies hebben een verband gevonden tussen de kwaliteit van evaluaties en benutting. Hierbij wordt opgemerkt dat dit verband van minder belang is dan het verband met aanbevelingen en communicatie. Van de zes studies was er één studie waarbij dit verband niet werd gevonden.

2.5.4 *Relevance*

Deze factor gaat in op het nut van de bevindingen en de mate waarin de bevindingen een handvat voor verbetering opleveren. Ten aanzien van deze factor zijn de conclusies niet eenduidig. Twee van de zes studies vinden geen relatie van belang, terwijl in twee andere studies wel een belangrijke relatie tussen relevantie van de informatie en benutting wordt vastgesteld.

2.5.5 *Credibility*

De mate waarin het gerapporteerde objectief is. De geloofwaardigheid van de beoordelaar in termen van het gehanteerde proces en de daarbij behorende activiteiten staat hier centraal. Van de vier studies die deze factor noemen, wordt in twee een sterke relatie met benutting gevonden, terwijl dit in de andere twee niet het geval is.

2.5.6 *Evaluator competence*

De persoonlijke bekwaamheid van de beoordelaar. Deze factor wordt in zes studies genoemd. In de meeste gevallen wordt verondersteld dat er een positief effect uitgaat van de bekwaamheid van de beoordelaar.

2.5.7 *Findings*

Ten aanzien van de resultaten van de evaluatie wordt niet alleen gezien wat voor resultaten zijn geboekt, maar ook de mate waarin deze overeenkomen met de verwachtingen en de waarde van de resultaten. Uit twee van de zes studies bleek deze factor van belang in relatie tot benutting, maar minder belangrijk als factoren als communicatie, timeliness en de kwaliteit van de evaluatie.

2.6 Findings – de aard van resultaten

In het bovenstaande overzicht is het relatieve belang van de factor ‘findings’ ten opzichte van de andere factoren uit de categorie ‘evaluation implementation’ verwoord. De centrale vraag van dit onderzoek gaat echter over de aard van de resultaten en de invloed hiervan op de benutting. Over de aard van de bevindingen is in de literatuur slechts weinig terug te vinden. Johnson noemt in zijn overzicht (Johnson 2009) twee elementen, die van belang zijn voor dit onderzoek:

- a. de waarde van de bevindingen voor de besluitvorming;
- b. positieve of negatieve resultaten.

In de twee meest recente artikelen die door Johnson worden aangevoerd ten behoeve van de classificering van de factor ‘findings’ als factor van belang ten opzichte van andere factoren, worden in deze artikelen ook bemerkingen gemaakt die relevant zijn voor het onderzoek. Vandaar dat deze artikelen hieronder nader zijn beschouwd.

2.6.1 Bober en Bartlett (2004)

Bober en Bartlett hebben in 2004 onderzoek gedaan naar de benutting van trainingsprogramma's in bedrijfsuniversiteiten. De studie was ontworpen om de toepasbaarheid van het raamwerk van Cousins en Leithwood (1986) te testen op benutting van evaluaties. Zij stellen daarbij dat – wil een evaluatie echt een formatief of summatief doel dienen – de resultaten moeten worden benut. Zij onderstrepen deze stelling met de uitspraak dat hoewel er een aantal componenten van invloed zijn op de evaluatie, de benutting van de evaluatieresultaten constant in de literatuur wordt genoemd als de meest belangrijke, maar dat deze tevens het meest over het hoofd wordt gezien (Piskurich, 1997; Robinson en Robinson, 1989; Wentling, 1980). Daarnaast verwijzen ze naar het werk van Stufflebaum (Stufflebaum, 2001), die opmerkt dat benutting georiënteerde evaluatie helpt de aandacht te sturen in de richting van resultaten die een impact hebben. *‘Information that assists others in making judgments about a program, service, organization, person or whatever else is being evaluated’* (Henry, 2001, p.1).

Ten aanzien van het belang van factoren constateren Bober en Bartlett ten eerste, dat factoren uit de groep ‘karakteristieken van evaluatie-implementatie’ van groter belang waren dan de factoren uit de groep ‘karakteristieken van besluitvorming of beleidskader’. Hoewel de factor ‘findings’ zich wel in de hoogste categorie bevindt, scoort hij in die categorie slechts als zevende factor van belang (Bober en Bartlett, 2004, p.377, tabel 2).

Relevant voor het onderzoek is de verwijzing in het artikel naar eerder onderzoek dat is uitgevoerd door Preskill (Preskill, 1984). Zij constateert in haar studie dat evaluatieresultaten zowel van instrumentele als van conceptuele aard kunnen zijn. Een veel groter percentage van benutting bleek echter van instrumentele aard. Preskill realiseerde zich dat het ontdekken van benutting van conceptuele aard veel lastiger is. Ze was er wel van overtuigd dat, indien ze betere toegang tot het hogere management had gehad, ze deze vorm van benutting had kunnen identificeren en bestuderen. De gevonden resultaten uit het onderzoek van Preskill komen overeen met de resultaten uit het onderzoek van Bober en Bartlett en worden zichtbaar gemaakt in tabel 1 (Bober en Bartlett, 2004, p.375, tabel 1).

2.6.2 Weiss, Murphy-Graham en Birkeland (2005)

Het onderzoek dat Weiss e.a. vanaf 2000 uitvoerden en in 2005 over rapporteerden, betrof een onderzoek naar het Drug Abuse Reduction Education oftewel het D.A.R.E-programma. Het beoogt kinderen te ontmoedigen drugs te gebruiken. Het programma was al vaak onderwerp van onderzoek

geweest, waarbij steeds de conclusie werd getrokken dat het programma niet effectief was. Ondanks deze uitkomsten bleek nog steeds 70 tot 80 % van de schooldistricten het programma uit te voeren. Dit geval van 'non-benutting' leek een interessant onderzoeksobject. Gedurende het onderzoek werd echter duidelijk dat zowel de publieke opinie als de organisatie die het programma subsidieerde, wel degelijk beïnvloed werden door de evaluatieresultaten. Schooldistricten dienden uiteindelijk te bewijzen dat het programma hielp in de reductie van druggebruik bij scholieren. Anders zouden zij de subsidie voor het programma verliezen. Als gevolg hiervan lieten diverse districten het programma vallen.

In deze studie wordt aangetoond dat resultaten van een evaluatie (de 'findings') een dominante factor waren in de benutting van de evaluatie. Daarnaast werd een tendens waargenomen dat er in toenemende mate sprake was van een dwingend karakter.

Relevant voor het onderzoek zijn ook de opmerkingen die Weiss e.a. maken ten aanzien van het steeds meer in zwang zijnde 'best practices' model. Hierbij worden programma's onderzocht met als doel de meest effectieve/efficiënte werkwijzen te identificeren. Weiss e.a. stellen dat deze vorm van evaluatie uitnodigt om druk uit te oefenen de 'best practices identified' breed toe te passen (Weiss e.a 2005. p. 26).

2.7 Typen benutting

De afgelopen veertig jaar is veel geschreven over de benutting van evaluatieresultaten. Er is vooral veel empirisch onderzoek naar gedaan tegen het einde van de jaren zeventig en begin jaren tachtig. Henry en Mark (2003) noemen deze periode de 'golden age' van onderzoek naar evaluatiebenutting. Dit onderzoek heeft ertoe geleid dat er ten aanzien van de benutting van evaluaties drie klassieke vormen van benutting worden onderscheiden: instrumentele, conceptuele en symbolische (of politieke). Deze vormen van benutting vindt men terug in het theoretisch model, dat is gebaseerd op de algemeen erkende en vaak gebruikte classificatie door Leviton en Hughes (1981). Daarnaast wordt procesgerelateerde benutting ('process use') van evaluaties genoemd, zoals aangegeven door Patton (1998). Deze vier vormen van benutting worden door Burke Johnson (1998) aangeduid als de 'meest voorkomende vormen van evaluatiebenutting'. Daarnaast zijn er diverse andere vormen van benutting gedefinieerd. Dal Santo, Goldberg, Choice en Austin (2002) hanteren de termen overtuigende benutting, methodologische benutting en indirecte benutting. Baltasar introduceert in 2009 'algemeen gebruik' als separate vorm van benutting.

2.7.1 Instrumentele benutting

Instrumentele benutting betekent het directe gebruik van resultaten van de evaluatie door beleidsmakers en andere spelers: '*Instrumental use refers to utilization that can be documented*' (Rich 1991, 333). Patton (1997) noemt deze vorm van benutting: '*intended use by intended users*'. Instrumentele benutting (de implementatie van de aanbevelingen) biedt voordelen als degenen, die verantwoordelijk zijn voor de maatregelen, de kosten van de evaluatie zelf dragen. Daarnaast is het nut van de evaluatie groter als die is gericht op de behoeften van de potentiële gebruikers. Weiss (1998) verbindt aan dit type evaluaties drie voorwaarden:

- a. de implicaties van de bevindingen zijn relatief niet-controversieel en veroorzaken geen frictie in de organisatie.
- b. de veranderingen passen binnen het bestaande programma en zijn relatief kleinschalig.
- c. de omgeving van het programma is relatief stabiel.

2.7.2 Conceptuele benutting

Het tweede type van benutting is conceptuele benutting van evaluatieresultaten. Conceptuele benutting vindt plaats wanneer beslissers hun meningen, attitudes of ideeën met betrekking tot bepaalde aspecten van het geëvalueerde programma veranderen als gevolg van een evaluatie. Carol Weiss (1977, p. 535) verklaart dat de conceptuele benutting van wetenschappelijke kennis bij beleidsuitvoering plaatsvindt door middel van de *'gradual sedimentation of insights, theories, concepts and ways of looking at the world'*. Het gaat hier om het begrip en het verkrijgen van nieuwe ideeën en inzichten. Men ontdekt sterke en zwakke punten en mogelijke richtingen voor actie (Weiss, 1998).

2.7.3 Symbolische (of politieke) benutting

Er is sprake van symbolische benutting indien beleidsmakers gebruik maken van evaluaties om hun perspectief te bevestigen en om legitimatie te krijgen voor zichzelf (Henry en Rog 1998). De evaluatie wordt gebruikt om te rechtvaardigen wat beleidsmakers hoe dan ook willen. Evaluatie kan in een dergelijk geval ondersteuning bieden aan het beleid op basis van intuïtie, beroepservaring, eigenbelang, organisatorische interesse, een zoektocht naar prestige, of mogelijk nog andere redenen die ingaan op beslissingen over het beleid en de praktijk. Weiss (2009) is positiever. Ze omschrijft deze vorm van benutting als: het mobiliseren van ondersteuning voor noodzakelijke veranderingen in het programma. Er wordt gebruik gemaakt van evaluatie als legitimatie. De evaluatie wordt een instrument van overreding/overtuiging. Baltasar (2009) suggereert op grond van de reacties van andere 'stakeholders' dat symbolische benutting toeneemt naarmate de afstand tussen beoordelaars en geëvalueerden groter wordt.

2.7.4 Andere vormen van benutting

Naast de drie klassieke – en in dit onderzoek gebruikte – vormen van benutting zijn er door de jaren heen ook andere vormen van benutting gedefinieerd. Voor de volledigheid wordt een aantal van deze vormen hieronder benoemd en kort toegelicht.

- a. Overtuigende benutting. Overtuigende benutting (persuasive utilization) vindt plaats wanneer evaluatieresultaten gebruikt worden om een bepaald standpunt te ondersteunen. Voor dit type benutting wordt verwezen naar het werk van Leviton en Hughes uit 1981.
- b. Methodologische benutting. Methodologische benutting vindt plaats als een beoordelaar een specifieke evaluatie 'tool' adopteert die of kan worden gebruikt als standaard test, of als een diagnostisch middel (Tripodi, Fellin en Meyer, 1983).
- c. Indirecte benutting. Er is sprake van indirecte benutting indien beoordelaars theorieën, modellen of procedures gebruiken, die weliswaar gebaseerd zijn op onderzoek, maar waar geen direct contact met de onderzoeksresultaten nodig is (Reid en Fortune, 1992).
- d. Algemene benutting. Algemene benutting is niet een andere onafhankelijke vorm van benutting, maar eerder een categorie die de beoordeling van het totale nut van de evaluatie omvat. 'Uit ervaringen van andere projecten blijkt, dat het vaak makkelijker is om het nut als geheel te beoordelen dan de beoordeling uit te splitsen naar de verschillende soorten van benutting' (Balthasar 2000).
- e. Imposed use. Weiss e.a (2005) introduceren de term 'imposed use' en omschrijven die als volgt: *'... where a higher level of government with funds to disburse demands specific action on lower operating levels based on evidence'* (Weiss e.a. 2005, p.25). Weiss e.a. geloven dat deze vorm van benutting steeds vaker zal worden gebruikt. Fondsen (in dit geval publieke fondsen) dienen verstandig te worden aangewend. Het feit dat controle plaatsvindt op grond van wetenschappelijk bewijs is een krachtig signaal. Weiss e.a. concluderen dat er in toenemende mate aandacht is – en druk bestaat – om wetenschappelijk bewijs te gebruiken en te benutten.

2.8 Samenvatting

In dit hoofdstuk is een relevant kader voor het onderzoek geschapen. Naast generieke begripsbepaling, is antwoord gegeven op de eerste vier deelvragen, zoals die in het eerste hoofdstuk zijn geformuleerd. Ten eerste is ingegaan op de vraag wat er wordt verstaan onder benutting. Hierbij is vastgesteld dat er van benutting sprake is indien de evaluatie een effect genereert op het geëvalueerde programma. Bovendien is antwoord gegeven op de vraag: 'wat wordt er benut' en de vraag: 'hoe wordt benut'. Deze laatste vraag beschrijft vooral kenmerken die van invloed zijn op het vermogen van een organisatie om te veranderen.

De tweede deelvraag is beantwoord door de verschillende factoren te beschrijven die benutting beïnvloeden. Als basis is het raamwerk van 'Cousins en Leithwood' gebruikt. Hierbij zijn twee categorieën onderscheiden: karakteristieken van evaluatie implementatie en karakteristieken van besluitvorming- of beleidskader. Dit raamwerk is aangevuld met het werk van Johnson. Naast de categorieën die worden gebruikt in het raamwerk wordt door Johnson nog een nieuwe categorie geïntroduceerd te weten 'stakeholder involvement'. Een tweede aanvulling van Johnson betreft het toevoegen van de factor 'evaluator competence' in de categorie 'evaluation implementation'. Als derde bouwsteen is Patton's 'process use' in het overzicht opgenomen.

Vervolgens is nader ingegaan op de factoren uit de categorie 'evaluation implementation'. In deze categorie bevindt zich ook de factor 'findings' of 'de aard van resultaten'. Deze factor is als antwoord op de derde deelvraag, nader uitgewerkt. Tot slot hebben de diverse vormen van benutting de revue gepasseerd. Hierbij zijn niet alleen de klassieke drie vormen van benutting (instrumenteel, conceptueel en symbolisch of politiek) uitgewerkt, maar wordt ook een overzicht geboden over andere vormen van benutting.

Het volgende hoofdstuk gaat in op de gehanteerde methodiek.

3 Methodologie

3.1 Inleiding

Waar in het vorige hoofdstuk aandacht is geschonken aan de theoretische basis, beoogt dit hoofdstuk antwoord te geven op de vraag hoe het onderzoek is ingericht. In het hoofdstuk worden onderwerpen aangestipt die voor een methodische onderbouwing zorg dragen. Zo passeren de keuze voor en het ontwerp van de gevalsstudie de revue. Daarnaast wordt ingegaan op de verschillende typen gevalsstudie. Vervolgens wordt aandacht besteed aan datacollectie, de principes die hierbij moeten worden gehanteerd en de verschillende analysetechnieken. Tot slot wordt ingegaan op de kwaliteitseisen waaraan het onderzoek moet voldoen.

Deze onderwerpen worden toegepast op de gekozen gevalsstudie en verklaren de gehanteerde onderzoeksmethode.

3.2 De gevalsstudie

Uitgangspunt bij de bepaling van de onderzoeksstrategie is het boek 'Case Study Research' van Robert K. Yin. In de onderzoekstrategie zullen dan ook voornamelijk kwalitatieve redeneringen en methoden een rol spelen. Daarom dient de centrale onderzoeksvraag zodanig te zijn dat hij voldoet aan de volgende eigenschappen (Yin 2009 p. 8 ev.):

- a. Het type onderzoeksvraag gaat in op het Hoe? En Waarom?
- b. Het vereist geen controle van de onderzoeker op de uitkomst van het resultaat en/of de gedragingen.
- c. Het onderwerp beschouwt de huidige situatie (in tegenstelling tot historische gebeurtenissen).

Voorts is kenmerkend voor gevalsstudies dat er sprake is van een eenheid, of slechts enkele eenheden, uit een verzameling die het domein van de probleemstelling vormt. Daarnaast betreft het een bestudering van de werkelijkheid – de natuurlijke omgeving – van een verschijnsel gedurende een specifieke periode. Er worden verschillende bronnen van informatie gebruikt om verklaringen van het te onderzoeken verschijnsel te vinden en deze zo mogelijk te veralgemenen. Deze verklaringen kunnen worden getoetst door de onderzochten zelf.

3.3 Ontwerp van een gevalsstudie

3.3.1 Componenten

Bij het ontwerp van de gevalsstudie gaat het er om dat er een logisch plan ontstaat. Een onderzoeksontwerp brengt een logisch verband aan tussen de te verzamelen data en de te trekken conclusies in relatie tot de onderzoeksvragen. Yin citeert een definitie van Nachmias & Nachmias (Yin 2009 p. 26): *'Guides the investigator in the process of collecting, analyzing and interpreting observations. It is a logical model of proof that allows the researcher to draw influences concerning casual relations among the variables under investigations'*.

Yin acht vijf componenten cruciaal voor een goed ontwerp (Yin 2009 p. 27 ev.):

1. de onderzoeksvraag;
2. een (probleem)stelling (indien nodig);
3. de analyse eenheid;

4. het verband van de resultaten (data) met de probleemstelling;
5. de criteria waarlangs de resultaten beoordeeld worden.

3.3.2 *Typen gevalsstudies*

Yin maakt (Yin, 2009 p. 46 ev.) een primair onderscheid in gevalsstudies tussen enkelvoudige (type 1 en 2) en meervoudige gevalsstudies (type 3 en 4).

Yin (Yin 2009 p. 60 ev.) adviseert om, indien dit mogelijk is, te kiezen voor een meervoudige gevalsstudie. Dit vanwege het kwetsbare karakter van enkelvoudige gevalsstudies en het feit dat meervoudige gevalsstudies voordelen ten aanzien van de analyse opleveren. Toch zijn in bepaalde omstandigheden enkelvoudige gevalsstudies een geschikte methode. Yin heeft de volgende argumenten:

- a. als het een kritieke kwestie in het testen van een goed geformuleerde theorie betreft;
- b. als het een extreme of unieke kwestie betreft;
- c. als het een representatieve of typische kwestie is voor het achterhalen van omstandigheden en condities van alledaagse situaties;
- d. als het een onthullende kwestie is die voorheen niet kon worden onderzocht;
- e. als het een langdurige kwestie is met een onderzoek op twee of meer momenten in de tijd.

	Single case design	Multiple case designs
Holistic (single level/unit of analysis)	Type 1	Type 2
Embedded (multiple levels/units of analysis)	Type 3	Type 4

Basic types of designs for case studies (Yin, 1994)

Het onderscheid tussen type 1 en type 2 is gelegen in het feit dat bij type 1 (holistic) slechts sprake is van 1 niveau/analyse-eenheid en bij type 2 er meerdere (embedded) niveaus/analyse-eenheden te onderscheiden zijn.

Kenmerkend voor meervoudige gevalsstudies is de herhalingslogica. Hierbij maakt Yin (Yin 2009 p. 54) een onderscheid tussen letterlijke en theoretische herhaling. Bij een letterlijke herhaling worden gelijke resultaten voorspeld, terwijl bij een theoretische herhaling tegengestelde (maar wel geanticipeerde) resultaten worden verwacht die theoretisch kunnen worden verklaard. Voorwaarde is dat in beide gevallen een goed theoretisch raamwerk moet worden ontwikkeld.

3.4 Datacollectie

3.4.1 Onderzoeksbronnen

Yin benoemt de zes belangrijkste onderzoeksbronnen. Hierbij tekent hij aan dat dit geen complete lijst is, maar dat meerdere bronnen beschikbaar zijn. Daarnaast merkt hij op dat er geen hiërarchie bestaat. Iedere bron heeft zijn eigen sterktes en zwaktes (Yin 2009 p. 102).

- a. Documentatie.
- b. Archieven.
- c. Interviews.
- d. Directe observatie.
- e. Participatieve observatie.
- f. Voorwerpen.

3.4.2 Drie principes bij datacollectie

Om de betrouwbaarheid en begripsvaliditeit van de verzamelde gegevens – en daarmee de uitkomst van het onderzoek – te waarborgen, dienen volgens Yin (Yin 2009 p. 114) drie principes in acht te worden genomen. Deze drie principes betreffen:

- a. het gebruik van meerdere bronnen ('triangulation');
- b. het creëren van de 'case study database';
- c. het handhaven van de 'chain of evidence'.

3.4.3 Vijf technieken voor analyse

Yin (Yin 2009 p. 136 ev.) beschrijft vijf technieken die kunnen worden toegepast bij het analyseren van gevalsstudies.

- a. Pattern matching: vergelijken van een empirisch gevonden patroon met een voorspeld patroon. Als deze patronen overeenstemmen, helpen de resultaten om de interne validiteit te verhogen.
- b. Explanation building: de uiteindelijke verklaring is het resultaat van een iteratief analyse proces.
- c. Time-series analysis: heeft betrekking op data die verzameld is op verschillende tijdstippen en is vergelijkbaar met kwantitatieve time-series analysis in experimenten en quasi-experimenten.
- d. Logic models: deze techniek verbindt empirisch gevonden uitkomsten met theoretisch beredeneerde uitkomsten (logic models zijn dan ook te vergelijken met pattern matching).
- e. Cross-case synthesis. Deze techniek is vooral geschikt voor de analyse van 'multiple' cases, waarbij wordt gezocht naar overeenkomsten in de uitkomsten.

3.5 Kwaliteitseisen

Het ontwerp moet voldoen aan een aantal kwaliteitseisen. Yin hanteert vier algemeen gangbare criteria om de kwaliteit van onderzoek te testen. Hij verwijst hierbij naar de omschrijving van L. Kidder & Judd, 1986 (Yin 2009 p. 40 ev.).

3.5.1 Begripsvaliditeit (construct validity)

Hiermee wordt bedoeld dat de interpretatie van de onderzoekgegevens zodanig moet zijn dat op legitieme wijze een verband kan worden gelegd tussen onderzoekresultaten en theoretische begrippen. Resultaten dienen passend te zijn en dekken op een adequate wijze de te meten theoretische begrippen. Bovendien moeten de juiste meetinstrumenten worden gebruikt (indiceren en operationaliseren). Yin geeft drie tips om de validiteit te optimaliseren:

- a. gebruik maken van meerdere bronnen;
- b. creëren van een keten in de bewijsvoering;
- c. beoordelen van de rapportage.

3.5.2 *Interne validiteit (internal validity)*

Bij de interne validiteit ligt het accent op de juistheid van de, in het conceptuele model veronderstelde, relatie tussen het onderzochte verschijnsel en zijn achtergronden. Men kan echter tevens zoeken naar een samenhangend interpretatiekader. Dit komt neer op het zoeken naar symptomen behorend bij een bepaalde hypothese. Deze techniek wordt ook wel pattern matching genoemd. Het zoeken naar een samenhangend interpretatiekader geeft geen garanties dat de onderzoeker het bij het juiste eind heeft. Echter, naarmate de onderzoeker er beter in slaagt zijn conceptuele model op te bouwen als een geheel van empirisch onderbouwde en logisch consistente uitspraken, wordt het aannemelijker dat deze uitspraken als een valide interpretatie van de situatie kan worden beschouwd. Yin geeft de volgende tips:

- a. gebruik pattern matching;
- b. gebruik explanation building;
- c. schenk aandacht aan strijdige verklaringen;
- d. gebruik logische modellen.

3.5.3 *Generaliseerbaarheid (external validity)*

Met de generaliseerbaarheid wordt bedoeld, de mate waarin een bredere reikwijdte aan de resultaten dan het onderzochte geval zelf kan worden toegekend. Om dit doel te bereiken dient een zorgvuldige selectie van de gevallen plaats te vinden. Met deze selectie wordt zoveel mogelijk geanticipeerd op argumenten, die de overtuigingskracht van de data in dit opzicht zouden kunnen ondergraven. Yin adviseert:

- a. gebruik van theoretische concepten in 'single-case studies';
- b. gebruik herhalingslogica in 'multiple case studies'.

3.5.4 *Betrouwbaarheid (reliability)*

De mate waarin een waarneming stabiel is bij verschillende metingen wordt de betrouwbaarheid genoemd. De empirische toetsing op betrouwbaarheid vereist dus de vaststelling of, en zo ja in hoeverre, de resultaten bij herhaling van het onderzoek op dezelfde (of gelijksoortige) onderzoekseenheden met dezelfde meetinstrumenten en dezelfde (of een soortgelijke) context overeenkomen. Door het relatief open karakter van de gehanteerde onderzoeksprocedures en de complexe structuur van de onderzochte situatie ontbreekt bij gevalsstudies in hoge mate de mogelijkheid om de betrouwbaarheid door middel van herhaling aan te tonen. Door de in het onderzoek gehanteerde procedures expliciet te maken en systematisch te werken wordt andere onderzoekers de mogelijkheid geboden om de stappen in de verzameling resultaten en het gebruik daarvan in de uiteindelijke conclusies te controleren. Ze moeten als het ware over de schouder van de onderzoeker kunnen meekijken. Om de betrouwbaarheid van een gevalsstudie te verhogen worden de volgende twee belangrijke maatregelen door Yin aanbevolen:

- a. de ontwikkeling van een gedegen uitgewerkt schriftelijk protocol voor de uitvoering van een gevalsstudie;
- b. het handhaven van een apart databestand voor het verzamelde materiaal.

3.6 **Keuzes voor het onderzoek verklaard**

3.6.1 *De keuze voor gevalsstudie*

Zoals reeds is aangegeven dienen gevalsstudies aan drie voorwaarden te voldoen. Ten eerste is de methode vooral geschikt voor 'hoe' en 'waarom' vragen (Yin, 2009). De centrale vraag van dit onderzoek voldoet aan die voorwaarde. De tweede voorwaarde is de mate van controle over de uitkomst van het resultaat en/of gedragingen. Invloed op de uitkomst van het onderzoek was niet

mogelijk. Dit betekent dat het onderzoek ook aan deze tweede voorwaarde voldoet. De derde voorwaarde betreft de focus op huidige versus historische gebeurtenissen. Gevalsstudie beschouwt de huidige situatie. Het onderzoek richt zich op een hedendaags verschijnsel en voldoet daarom ook aan de derde voorwaarde.

Daarbij voldoet het onderzoek aan de vier eigenschappen die kenmerkend zijn voor gevalsstudies. Het onderzoek richt zich op cases die een representatieve uitsnede vormen van de gehele 'lessons learned' database. Daarnaast beslaan de cases een specifieke periode (eind 2010 – eind 2011). Verder zijn verschillende databronnen gebruikt om de uitkomsten van het onderzoek te verklaren, waarbij deze uitkomsten zijn voorgelegd en getoetst door de onderzochten zelf.

Tot slot kan de keuze voor de gevalsstudie worden afgezet tegen andere onderzoeksstrategieën, zoals genoemd door Yin (2009 p.8 figuur 1.1). Ook dan blijkt de gevalsstudie de beste keuze. De centrale vraag impliceert dat het onderzoek zich richt op een hedendaags verschijnsel. Dit betekent dat een keuze voor historisch onderzoek en archiefanalyse (in zijn pure vorm) niet logisch is. Het gebrek aan controle over uitkomst en/of gedrag impliceert dat ook het experiment als onderzoeksstrategie afvalt. Resteert de keuze tussen de gevalsstudie of een survey. Het voornaamste bezwaar tegen de uitvoering van een survey is gelegen in het feit dat bij een survey de context van de onderzochte eenheid of eenheden uit het oog wordt verloren. Deze context is – gezien de centrale vraag – cruciaal bij het uitvoeren van dit onderzoek.

3.6.2 *Het ontwerp van de studie*

Het ontwerp van de studie volgt de componenten die door Yin noodzakelijk worden geacht voor een goed ontwerp. Zo kent het onderzoek een centrale onderzoeksvraag, die wordt beantwoord aan de hand van een aantal deelvragen. De uitkomsten van het onderzoek worden beoordeeld op grond van geoperationaliseerde criteria.

Ten aanzien van het type gevalsstudie is gekozen voor een meervoudige gevalsstudie. De voorwaarden die Yin stelt voor het uitvoeren van een enkelvoudige gevalsstudie (zie de argumenten die worden genoemd in para 3.3.2.) passen niet bij de voor dit onderzoek geformuleerde onderzoeksvraag. Er is daarom gekozen voor een ontwerp, waarbij gelijkgerichte (gelijkvormige) cases in de periode van eind 2010 tot eind 2011 zijn onderzocht. Door de gelijkvormigheid van deze cases (het betreffen 3 opeenvolgende F-16 detachementen) is het mogelijk geweest bij de analyse van de evaluatieresultaten gebruik te maken van herhalingslogica. Met deze keuze is tevens een balans gevonden waarbij enerzijds voldoende tijd is gegeven om te bezien op welke wijze benutting heeft plaatsgevonden, anderzijds zijn de cases voldoende actueel, waarmee 'verlies van geheugen' wordt voorkomen.

3.6.3 *Datacollectie*

Voor het onderzoek is vastgehouden aan de drie principes van datacollectie, zoals die door Yin zijn geformuleerd. Er zijn drie onderzoekbronnen gebruikt. Ten eerste de 'lessons learned' database. Deze database bevat alle resultaten van evaluaties die hebben plaatsgevonden in het kader van het 'lessons learned' proces. Daarnaast is in deze database opgenomen of, en zo ja in hoeverre, de resultaten ook daadwerkelijk hebben geleid tot verandering c.q. zijn geïmplementeerd. Een tweede bron van informatie werd gevormd door de evaluatieverslagen zelf. Deze verslagen (en zo mogelijk ander achtergrondmateriaal) zijn gebruikt om een diepere analyse uit te voeren ten aanzien van de gevonden resultaten. Door de bestudering van de verslagen konden de evaluatieresultaten in hun

context worden geplaatst. De derde bron van informatie betrof de gevoerde gesprekken. Het gebruik van genoemde bronnen maakte triangulatie mogelijk.

De 'lessons learned' database kon voor het onderzoek zodanig worden bewerkt, dat er een 'case study database' ontstond, waarmee voldaan werd aan het tweede, door Yin genoemde, principe.

De gevoerde gesprekken waren naast een belangrijke bron van informatie – een nog scherper beeld op de vraag of, en zo ja in hoeverre, benutting heeft plaatsgevonden – gebruikt als toetsingsmechanisme voor de geanalyseerde data. Hiermee is bereikt dat de 'chain of evidence' werd opgebouwd en gehandhaafd.

3.6.4 *Begripsvaliditeit (construct validity) en Interne validiteit (internal validity).*

Om de begripsvaliditeit van het onderzoek te optimaliseren, is – zoals hierboven beschreven – gebruik gemaakt van verschillende bronnen. Voor wat betreft de onderzochte data zijn niet alleen de evaluatieverslagen zelf gebruikt, maar is ook de achterliggende documentatie geraadpleegd. Daarnaast zijn met diverse betrokkenen gesprekken gevoerd. Als informatiebron op zich, maar ook om verklaringen te vinden voor de gevonden resultaten. Tegenstrijdigheden – hetzij in de verzamelde data of verklaringen, hetzij ten opzichte van de geraadpleegde literatuur – zijn nader geanalyseerd en tijdens de gesprekken voorgelegd. Ten einde de begripsvaliditeit verder te verbeteren zijn ook de uitkomsten van het onderzoek ter beoordeling voorgelegd aan de geïnterviewden.

Bovengenoemde werkwijze heeft niet alleen de begripsvaliditeit verbeterd, maar is ook de interne validiteit ten goede gekomen. Er is tijdens het onderzoek getracht om van de conclusies en uitspraken van het onderzoek een zo consistent en samenhangend mogelijk geheel te maken. Hierbij is gebruik gemaakt van pattern matching en explanation building. Van de analyse is een iteratief proces gemaakt en zijn verbanden gelegd tussen de verschillende gebruikte bronnen.

3.6.5 *Generaliseerbaarheid (external validity).*

Aan de resultaten van het onderzoek kan een bredere reikwijdte worden toegekend dan aan het geval zelf. Hierbij dienen echter wel de beperkingen van het onderzoek voor ogen te worden gehouden. In het onderzoek is gebruik gemaakt van herhalingslogica, vooral voor wat betreft de data uit de verschillende evaluatieverslagen.

3.6.6 *Betrouwbaarheid (reliability).*

De in het onderzoek gebruikte metingen zijn zo consequent mogelijk toegepast en zijn per onderzoeksvraag geoperationaliseerd (zie hoofdstuk 4). Hiermee is getracht een inzichtelijk protocol te creëren, waardoor het mogelijk is de resultaten van het onderzoek te controleren.

3.7 Samenvatting

In dit hoofdstuk is de keuze voor het uitvoeren van een gevalsstudie verklaard. Hierbij is vastgesteld dat het onderzoek voldoet aan de voorwaarden die Yin stelt aan het uitvoeren van een gevalsstudie. Daarnaast zijn in het hoofdstuk een viertal eigenschappen, die kenmerkend zijn voor gevalsstudies, besproken. Ook aan deze eigenschappen voldoet het onderzoek.

Het ontwerp van de studie volgt de componenten die door Yin noodzakelijk worden geacht voor een goed ontwerp. Ten aanzien van het type gevalsstudie is gekozen voor een meervoudige gevalsstudie. Er is gekozen voor een ontwerp, waarbij gelijkgerichte (gelijkvormige) cases in de periode van eind

2010 tot eind 2011 zijn onderzocht. Met de keuze voor deze cases is tevens een balans gevonden waarbij enerzijds voldoende tijd is gegeven om te bezien op welke wijze benutting heeft plaatsgevonden, anderzijds zijn de cases voldoende actueel, waardoor 'geheugenverlies' is voorkomen.

Voor het verzamelen van gegevens zijn een drietal onderzoekbronnen gebruikt:

- a. de 'lessons learned' database (deze is omgevormd tot een 'case study database');
- b. de evaluatieverslagen (deze zijn gebruikt voor het in context plaatsen van resultaten);
- c. gesprekken met direct betrokkenen (ter vergroting van het inzicht en ter toetsing verklaringen c.q. nader onderzoek bij tegenstrijdigheden).

Ten aanzien van de kwaliteit van het onderzoek is inzicht gegeven in de wijze waarop de kwaliteit is geborgd. Het gebruik van meerdere bronnen en de controlefunctie die is ingebracht tijdens de gesprekken met de verschillende betrokkenen, hebben bijgedragen aan zowel de begripsvaliditeit als de interne kwaliteit van het onderzoek. Door gebruik te maken van pattern matching en explanation building is getracht van de conclusies en uitspraken een zo consistent en samenhangend mogelijk geheel te maken. Aan de resultaten van het onderzoek kan een bredere reikwijdte (generaliseerbaarheid) worden toegekend, mede door bij de analyse van de evaluatieresultaten gebruik te maken van herhalingslogica. Hierbij dienen echter wel de beperkingen van het onderzoek voor ogen te worden gehouden. Tot slot is aandacht geschonken aan de betrouwbaarheid van het onderzoek. Beoordelingen zijn zo consequent mogelijk toegepast op basis van een inzichtelijk protocol, waardoor het mogelijk is de resultaten van het onderzoek te controleren.

Het volgende hoofdstuk gaat in op de resultaten van het onderzoek.

4 Resultaten van het onderzoek

4.1 Inleiding

In de voorgaande hoofdstukken is het kader voor het onderzoek geschapen. Dit hoofdstuk gaat in op de resultaten van het onderzoek zelf. Aan de hand van de – in het eerste hoofdstuk geformuleerde – deelvragen wordt toegewerkt naar de beantwoording van de centrale onderzoeksvraag. Daarnaast zijn in dit hoofdstuk de gehanteerde begrippen geduid en – waar nodig – geoperationaliseerd.

4.2 In hoeverre worden de resultaten die voortkomen uit een evaluatie ‘geleerde lessen’?

Deze vraag beoogt een overzicht te geven van de bereikte evaluatieresultaten. Hierbij is gebruik gemaakt van de ‘lessons learned’ database. Het overzicht biedt inzicht in de mate waarin ‘observations’ leiden tot geleerde lessen (‘lessons learned/best practices’). Daarnaast wordt in het overzicht een vergelijking gemaakt tussen de gehele database en de geselecteerde cases. Hiermee is bezien of/in hoeverre de geselecteerde cases representatief zijn. Tot slot is nagegaan of de geselecteerde cases voldoende ‘rijp’ waren, waardoor benutting plaats kon vinden (zie bijlage c).

Voor het lezen van het overzicht worden de termen uit de eerste kolom nader verklaard. Hiervoor wordt gebruik gemaakt van de definities zoals die door de NAVO worden gehanteerd⁴:

- a. **‘Observation’.** An Observation is *‘a comment based on something someone has heard, seen or noticed that has been identified and documented as an issue for improvement or a potential best practice’* (zie voor nadere informatie bijlage a).
- b. **Lesson Identified.** A Lesson Identified is *‘a mature observation with a determined root cause of the observed issue and a recommended remedial action and action body, which has been developed and proposed to the appropriate authority.’*
- c. **Best practice.** A Best Practice is *‘a technique, process or methodology that contributes to the improved performance of an organization and has been identified as a ‘best way of operating’ in a particular area as compared to other good practice(s). Ideally, a Best Practice should be adaptive, replicable and immediately useable’.*
- d. **Lessons Learned.** A Lesson Learned is *‘an improved capability or increased performance confirmed by validation when necessary resulting from the implementation of one or more remedial actions for a lesson identified.’*
- e. **Incidenten en afgedaan.** De termen ‘incidenten’ en ‘afgedaan’ komen niet voor in het ‘lessons learned’ proces. Deze termen zijn in de ‘lessons learned’ database opgenomen ter ondersteuning van de ‘lessons learned’ functionaris (zie bijlage c, vraag 6 en 7). De redenen voor het categoriseren van resultaten als ‘incident’ of ‘afgedaan’ zijn verschillend en voornamelijk van intern organisatorische aard. De categorie ‘incidenten’ is ingebracht om resultaten die eenmalig lijken en geen verbeteractie in zich lijken te dragen, toch te bewaren voor het geval herhaling van het resultaat op een later tijdstip plaatsvindt. Indien een dergelijk of soortgelijk incident nogmaals wordt genoemd, wordt dit opnieuw beoordeeld en – indien daar aanleiding toe bestaat – opnieuw geclassificeerd. De categorie ‘afgedaan’ is ingebracht om resultaten te parkeren die niet zullen leiden tot aanpassingen in werkwijze (bijvoorbeeld omdat het resultaat zeer situatie- en plaatsgebonden is waardoor herhaling onmogelijk). Daarnaast worden beide categorieën gebruikt om vervuiling van de database te voorkomen.

⁴ Bi-SC Command Directive (Bi-SCD) 80-6 Lessons Learned, 06 July 2011.

	Totale database	ATF 15	ATF 16	ATF 17	Totaal ATF (15/16/17)
OBSERVATIONS	82	1	7	5	13
INCIDENTEN	44	9	8	2	19
LESSONS IDENTIFIED	113	31	10	5	46
BEST PRACTICES	19	-	5	-	5
LESSONS LEARNED	36	4	3	-	7
AFGEDAAN	21	13	-	1	14
TOTAAL AANTAL RESULTATEN	315	58	33	13	104

Toelichting: De tabel bevat de data zoals die zijn opgenomen in de database. In de eerste kolom zijn de totalen van alle evaluaties opgenomen. Kolom twee, drie en vier bevatten de resultaten van de gekozen cases. Kolom vijf bevat de som van resultaten van de drie cases.

Om te komen tot een zuivere beoordeling, is de ordening van de data op twee plaatsen aangepast. Ten eerste zijn de drie cases gelijkvormig gemaakt. De resultaten van het detachement 15 zijn aangepast omdat dit detachement een andere samenstelling had (het beschikte naast F-16 gevechtswliegtuigen ook over een helikoptereenheid) dan de detachementen 16 en 17. Dit was eenvoudig te realiseren, aangezien het in de database zichtbaar is welke resultaten afkomstig waren van de helikoptereenheid (zie bijlage c, vraag 8). De tweede aanpassing betrof de indeling van de rijen. De categorieën 'incidenten' en 'afgedaan' komen niet voor in het 'lessons learned' systematiek. Daarom zijn de resultaten uit deze categorieën nader onderzocht en heeft in samenwerking met het betrokken 'lessons learned' functionaris herindeling plaatsgevonden (bijlage c, vraag 9 en eindnoot bijlage c). De resultaten uit deze categorieën zijn als volgt ingedeeld:

- a. **Incidenten.** Voor ATF 15 (na eerste opschoning) bleken de overgebleven drie resultaten, die als 'incident' waren gemerkt, zeer klein en eenmalig van aard. Deze 'incidenten' zijn uit de lijst verwijderd. Voor ATF 16 waren acht resultaten als 'incident' gemerkt. Van deze acht waren vier resultaten zeer klein en eenmalig van aard. Twee resultaten waren de bevestiging van een goed functionerende situatie, waardoor geen verdere actie nodig was. Deze 'incidenten' zijn uit de lijst verwijderd. In twee gevallen kon geen verbeterpunt voor de eigen organisatie worden geformuleerd. Ook deze twee resultaten zijn uit de lijst verwijderd. Van de twee resultaten die voor ATF 17 als 'incident' waren beoordeeld, viel één resultaat (ver) buiten het competentiegebied van de eigen organisatie en was eenmalig van aard. Dit resultaat is uit de lijst verwijderd. Voor het andere resultaat (ATF-17/005) bleek de markering als 'incident' niet correct. Dit resultaat is aangepast en opgenomen als 'lesson identified' (zie eindnoot bijlage c).
- b. **Afgedaan.** Voor ATF 15 (na eerste opschoning) zijn de overgebleven drie resultaten beoordeeld als 'afgedaan'. In het overzicht zijn deze resultaten (ATF-15/ 011, ATF-15/014 en ATF-15/017) opgenomen als 'lessons learned'. Eén resultaat is voor ATF 17 als 'afgedaan' aangemerkt. Ook dit resultaat (ATF-17/004) is opgenomen als 'lessons learned' (zie eindnoot bijlage c).

De aanpassingen leiden tot het volgende, opgeschoonde overzicht.

	Totale database	ATF 15	ATF 16	ATF 17	Totaal ATF (15/16/17)
OBSERVATIONS	82	1	9	3	13
LESSONS IDENTIFIED	113	11	10	8	29
BEST PRACTICES	19	-	8	-	8
LESSONS LEARNED	36	4	1	1	6
TOTAAL AANTAL RESULTATEN	250	16	28	12	56

4.2.1 Analyse

Zowel het gelijkvormig maken van de drie cases als het herschikken van de categorieën 'incidenten' en 'afgedaan' is naar het oordeel van de 'lessons learned' functionaris correct uitgevoerd (bijlage c vraag 8). Hoewel van beperkte waarde lijkt het geheel van de resultaten van de cases een soortgelijke verdeling op te leveren als de verdeling van de gehele database. De onderlinge resultaten van de cases verschillen zeer. Deze onderlinge verschillen zijn echter goed verklaarbaar. Ten eerste wordt het 'vrije' karakter van de evaluatievorm genoemd. De geëvalueerde commandant heeft de ruimte zelf te kiezen voor de opgebrachte resultaten en te behandelen thema's. Bij die keuze speelt de achtergrond van de geëvalueerde commandant een belangrijke rol. Van invloed zijn het vermogen en de kennis om (vanuit eerdere ervaringen) een tekortkoming in een bredere context te plaatsen, zich bewust te zijn van mogelijk achterliggende problematiek en de haalbaarheid om een eventuele oplossing te voorzien. Verder wordt de aandacht van de geëvalueerde commandant voor het aan te leveren verslag genoemd, waarbij de betrokkenheid van zijn staf bij de totstandkoming van het verslag en de tijdigheid van aanlevering een rol spelen (zie bijlage c vraag 10).

De keuze voor de gekozen cases in de periode van eind 2010 tot eind 2011 wordt juist geacht. Naar het oordeel van de 'lessons learned' organisatie is een goede balans gevonden, die enerzijds voldoende tijd geeft om het proces te doorlopen en anderzijds voldoende actueel is, waarmee 'verlies van geheugen' wordt voorkomen (zie bijlage c vraag 4).

Uit het overzicht blijkt dat van de 56 resultaten nog steeds 13 resultaten zijn gemerkt als 'observation'. Nadere beschouwing leert dat van de 13 'observations' er 5 'observations' zijn die niet binnen de eigen (Luchtmacht) organisatie kunnen worden opgepakt. Deze zijn doorspeeld naar het centrale krijgsmachtniveau voor verdere behandeling. Drie 'observations' zijn in een later stadium alsnog aangemerkt als 'lesson identified'. Verder bleken vijf 'observations' niet meer relevant. Reden hiervoor waren veranderende omstandigheden zoals een wijziging in proces of procedure, anders dan door het 'lessons learned' proces geïnitieerd (3) en doordat de 'observations' minimaal van aard bleken en er onvoldoende grond was om een verbeterslag te starten (2) (zie bijlage c vraag 11).

Van de 56 zijn 29 resultaten gecategoriseerd als 'lessons identified'. Als voornaamste redenen voor het merken als 'lessons identified' en niet als 'lessons learned'/'best practice' worden twee oorzaken genoemd. Ten eerste worden veel bemerkingen ook 'los' van het 'lessons learned' proces teruggekoppeld. De behandeling en verbeterslag vinden dan plaats buiten het zicht van de 'lessons learned' organisatie. Ten tweede gaat het om zaken die op centraal krijgsmachtniveau moeten

worden opgepakt (bijvoorbeeld in overleg met de Landmacht). Ook dan is niet altijd te achterhalen of, en op welke wijze het resultaat is verwerkt en geborgd (zie bijlage c vraag 12).

Naast genoemde inhoudelijke oorzaken blijken ook organisatorische factoren oorzaak voor haperingen in het proces (zie bijlage d, vraag 2 en bijlage c, vraag 11). Zo wordt capaciteitsgebrek (de capaciteit die nodig is alle resultaten te volgen en juist te categoriseren) binnen de 'lessons learned' organisatie genoemd. Daarnaast lijkt het belang van het 'lessons learned' proces niet altijd voldoende onderkend bij andere organisatie-elementen. Er wordt geconstateerd dat de wijze waarop de borging van het verbeterpunt plaatsvindt, niet consequent wordt teruggekoppeld naar de 'lessons learned' organisatie. Hierdoor blijft onzekerheid bestaan over de vraag of de 'les' daadwerkelijk is geleerd.

4.2.2 Subconclusie

In deze paragraaf is een overzicht geboden van de data zoals die zijn aangetroffen in de 'lessons learned' database. Vervolgens is in stappen verklaard hoe dit overzicht werd 'opgeschoond' voor dit onderzoek. Naar het oordeel van de 'lessons learned' organisatie gebeurde dit correct. Geconcludeerd kan worden dat de onderzochte cases (ondanks grote onderlinge verschillen) een soortgelijke verdeling volgen als de gehele database. Hieruit kan alleen met terughoudendheid – en met oog voor het beperkte belang – worden geconstateerd dat de gekozen cases representatief zijn voor het geheel. Verder bleken de gekozen cases qua timing voldoende tijd bieden om het proces te doorlopen, maar toch voldoende actueel te zijn, waarmee 'verlies van geheugen' is voorkomen.

4.3 In hoeverre is de 'aard van de resultaten' van invloed op de mate van de voortgang in het proces?

Met deze vraag is nagegaan of er een onderscheid bestond tussen de 'aard van de resultaten' en de voortgang in het 'lessons learned' proces. Voordat antwoord op deze vraag wordt gegeven is hieronder eerst het begrip 'aard van de resultaten' verder uitgewerkt en geoperationaliseerd.

4.3.1 Aard van de resultaten

In het onderzoek zijn twee aspecten gebruikt om de 'aard van de resultaten' te beschouwen. Ten eerste is een onderscheid gemaakt tussen positieve en negatieve resultaten. Dit onderscheid past naadloos op dit type van evaluatie. Niet alleen negatieve resultaten zijn van belang, maar er is nadrukkelijk ook oog voor positieve resultaten: de z.g. 'best practices'. Negatieve resultaten kenmerken zich door een tekortkoming in een werkwijze, proces, procedure of omstandigheid. Positieve resultaten wijzen vooral op een (vaak nieuwe) werkwijze, proces of procedure die goed heeft gewerkt en waarvan wordt voorgesteld deze breed toe te passen.

Het tweede aspect betreft de complexiteit van de implementatie. Het operationaliseren van dit aspect is minder eenvoudig. In het onderzoek worden resultaten die 'minder complex' en 'meer complex' onderscheiden. De indeling is uitgevoerd aan de hand van drie criteria (zie bijlage c vraag 14):

- a. De mate waarin het resultaat invloed heeft op verschillende facetten van de (binnen het 'lessons learned' proces) gehanteerde DOTMLPFI-factoren⁵. Eenvoudige problemen beperken zich tot een enkele DOTMLPFI-factor. Complexere problemen omvatten meerdere of zelfs alle DOTMLPFI-factoren, waarbij ook een onderlinge interactie tussen deze factoren een rol kan spelen.

⁵ D – Doctrine, O – Organization, T – Training, M – Material, L – Leadership, P – Personnel, F – Facilities, I – Interoperability

- b. Het tweede criterium betreft de absolute impact van de implementatie (in termen van geld, organisatorische verandering, enz.).
- c. Complexiteit die ontstaat door het hanteren van een 'gelijkheidsbeginsel'. De (beleids-) verandering moet passen in een groter geheel, breder toepasbaar zijn, terwijl de continuïteit in beleid moet worden gehandhaafd.

4.3.2 Analyse

Om de bovenstaande vraag te beantwoorden zijn drie tabellen opgesteld. In deze tabellen is per case een overzicht van de resultaten gemaakt. Tijdens gesprekken met de 'lessons learned' functionaris (zie eindnoot bijlage c) zijn alle resultaten doorgenomen en beoordeeld. Deze tabellen zijn opgenomen als bijlage B. Hierbij is de vraag opgesplitst in vier deelvragen:

- a. In hoeverre leiden positieve resultaten tot 'best practices'?
- b. In hoeverre leiden negatieve resultaten tot 'lessons learned'?
- c. In hoeverre leiden 'minder complexe' resultaten tot 'lessons learned'/'best practices'?
- d. In hoeverre leiden 'meer complexe' resultaten tot 'lessons learned'/'best practices'?

Uit de tabellen valt af te leiden dat van de 56 onderzochte resultaten er 15 als positief te duiden zijn. Hiervan zijn 8 aangemerkt als 'best practice'. Dit is een percentage van ongeveer 50%. Van de negatieve resultaten bleken 41 negatief. Van deze 41 resultaten zijn 6 gemerkt als 'lessons learned', een aanmerkelijk lager percentage (14,6%).

Indien gekeken wordt naar de implementatiecomplexiteit ontstaat een eenzelfde beeld. Van de 56 resultaten zijn er 32 geclassificeerd als 'minder complex'. Hiervan zijn 12 aangemerkt als 'lessons learned'/'best practice' (4 als 'lessons learned' en 8 als 'best practice'). Dit is een percentage van 37,5 %. Als 'meer complex' zijn 24 resultaten gemerkt. Hiervan zijn 2 resultaten gemerkt als 'lessons learned'/'best practice' geclassificeerd (8,3%).

Uit het bovenstaande kan worden afgeleid dat positieve resultaten die minder complex zijn, sneller en vaker leiden tot een geborgde verbetering. Als verklaring voor dit gegeven wordt niet alleen het feit genoemd dat het resultaat 'minder complex' is. Het betekent vaak ook dat het uitgezonden detachement de implementatie zelf al vorm heeft gegeven (men draagt in veel gevallen zelf al een oplossing aan), waardoor de borging van het resultaat eenvoudig kan worden gerealiseerd (zie bijlage c, vraag 16).

4.3.3 Subconclusie

Op grond van bovenstaande gegevens kan worden geconcludeerd dat er een significant verschil bestaat tussen positieve en negatieve resultaten. Positieve resultaten leiden in meer dan 50% van de gevallen tot 'best practices'. Bij negatieve resultaten is dit percentage aanmerkelijk lager. Daarnaast kan worden vastgesteld dat minder complexe resultaten eerder leiden tot geleerde en geborgde lessen.

4.4 Op welke wijze zijn resultaten die voortkomen uit het 'lessons learned' proces benut?

Hoewel het merken van een resultaat als 'lessons learned' c.q. een 'best practice' een afgerond geheel impliceert, wordt met deze vraag vooral bezien of er van benutting daadwerkelijk sprake is. En zo ja, welke vorm van benutting. De resultaten van het onderzoek zijn beschouwd aan de hand van de 'klassieke' drie vormen van benutting: instrumentele, conceptuele en symbolische (of politieke) benutting. Tijdens de gesprekken werden voor de bepaling van het type benutting de volgende, vereenvoudigde criteria gehanteerd:

- a. instrumentele benutting impliceert enige vorm van een beslissing/besluit;
- b. conceptuele benutting genereert ideeën en begrip (en daarmee discussie);
- c. symbolische (of politieke) benutting leidt tot ondersteuning van argumenten of rechtvaardiging van (geen) actie.

Voor de beantwoording van de vraag zijn de 14 resultaten ('best practices' en 'lessons learned') onderverdeeld – conform de onderzoekscriteria – in vier groepen:

Resultaten	
Positief/Minder complex	8
Positief/Meer complex	0
Negatief/Minder complex	4
Negatief/Meer complex	2
Totaal	14

4.4.1 Analyse

Ten aanzien van de eerste groep – positief en minder complex – bleek het in alle gevallen slechts te gaan om kleine wijzigingen in proces of procedure. De aanpassing kon door de eenheid zelf worden geïmplementeerd en door de 'lessons learned' organisatie geborgd, zonder ingrijpende gevolgen voor (of betrokkenheid van) andere delen van de organisatie. Hierdoor kon het proces betrekkelijk snel worden doorlopen. Er was sprake van instrumentele benutting (zie bijlage e, vraag 1).

Voor de tweede groep – positief en meer complex – zijn geen resultaten geïdentificeerd.

De derde groep – negatief en minder complex – bevat vier resultaten. Drie van deze resultaten bleken eenvoudige aanpassingen in proces, werkwijze of procedure. Voor deze drie resultaten kon worden vastgesteld dat er sprake was van instrumentele benutting. Eén resultaat (ATF 15/011) heeft geleid tot discussie, waarbij de afweging tussen veiligheid en kosten centraal stond (zie bijlage e, vraag 1). Uiteindelijk is gekozen voor de oplossing conform de aanbeveling in het evaluatieverslag (instrumentele benutting). Toch lijkt er in dit geval ook sprake te zijn geweest van conceptuele benutting (fundamentele discussie over hoeveel veiligheid mag kosten, waarbij de beslissers hun mening en attitude hebben veranderd als gevolg van het evaluatieresultaat) en zelfs van symbolische benutting (veiligheid als argument op zich) (zie bijlage e, vraag 1).

De laatste groep bevat twee resultaten. Deze resultaten – negatief en meer complex – betroffen een tekortkoming in afstemming en aansturing door verschillende managementlagen (ATF 15/005) richting de uitgezonden eenheid, waardoor een herdefinitie en herbevestiging van rollen, taken, verantwoordelijkheden en bevoegdheden noodzakelijk was. Het tweede resultaat betrof een tekortkoming in de logistieke keten (ATF 17/004). In dit geval is de geconstateerde tekortkoming opgelost door verbeteringen in systemen en middelen, en het inbrengen van additioneel (en beter gekwalificeerd) personeel. Hoewel de implementatie van de verbeteracties gecompliceerd bleek, werden voor beide resultaten de tekortkomingen onderkend. De benutting van deze resultaten lijkt dan ook het best te kwalificeren als instrumenteel (zie bijlage e, vraag 1).

4.4.2 Subconclusie

Op grond van bovenstaande analyse kan worden geconstateerd, dat alle – conform het 'lessons learned' proces – geboekte 14 resultaten instrumenteel van aard zijn. Slechts in één geval lijkt er ook

sprake van andere vormen van benutting. De aard van de resultaten lijkt in deze 14 gevallen geen factor van invloed.

4.5 Zijn er resultaten waarvan benutting niet als zodanig is (h)erkend?

Om deze vraag te beantwoorden zijn alle resultaten uit de geselecteerde cases, die – vooralsnog – niet hebben geleid tot een ‘lessons learned’ of ‘best practice’, geanalyseerd. Vervolgens is bezien of er wellicht toch sprake is geweest van enige vorm van benutting. Ook hier is het onderscheid gemaakt tussen positieve en negatieve resultaten enerzijds en complexe en minder complexe resultaten anderzijds. Daarnaast is een onderscheid gemaakt op grond van fase waarin een bepaald resultaat zich bevindt. Eerst is gekeken naar de resultaten die (nog steeds) geclassificeerd zijn als ‘observation’. Vervolgens zijn resultaten die zijn gemerkt als ‘lessons identified’ beschouwd.

4.5.1 Observations

Er zijn 13 resultaten geclassificeerd als ‘observation’. Een weergave hiervan is verwerkt in onderstaande tabel:

‘Observations’	
Positief/Minder complex	3
Positief/Meer complex	0
Negatief/Minder complex	4
Negatief/Meer complex	6
Totaal	13

4.5.2 Analyse

De positieve en minder complexe resultaten bleken alle drie aanbevelingen te zijn om (verschillende groepen) personeel ter plaatse extra te trainen in specifieke taken. Follow-up acties hebben buiten het zicht van de ‘lessons learned’ organisatie plaatsgevonden, maar waren in termen van benutting instrumenteel van aard (zie bijlage e, vraag 2).

Van vier resultaten die zijn gemerkt als negatief en minder complex, waren twee resultaten van personele aard, 1 resultaat betrof een logistieke en 1 resultaat een operationele tekortkoming. Alle vier resultaten hadden een incidenteel karakter. Gebrek aan terugkoppeling van de correctie bleek in alle gevallen de reden van de classificatie ‘observation’. De benutting van deze resultaten kon gekwalificeerd worden als instrumenteel (zie bijlage e, vraag 2).

Voor wat betreft de zes resultaten, die als negatief en complex zijn geclassificeerd, zijn twee resultaten om verschillende redenen onterecht als zodanig gemerkt. Eén resultaat betreft nagenoeg een herhaling van een al geleerde les en had moeten worden verwijderd (ATF 16/009). Het tweede resultaat was inmiddels opgepakt, maar niet verwerkt in de database. Van de vier andere resultaten is vastgesteld dat een eventuele actienemer zich buiten de Luchtmachtorganisatie bevindt. Van alle vier de resultaten kan echter ook worden vastgesteld dat – zonder dat dit heeft geleid tot een directe ‘oplossing’ – er wel degelijk sprake is geweest van benutting. Aandacht voor de opgebrachte resultaten heeft geleid tot een scherper inzicht en het belang ervan benadrukt. In 1 geval is een ‘sense of urgency’ bewerkstelligd. De benutting van deze laatste vier resultaten kan worden gekenmerkt als conceptuele benutting (zie bijlage e, vraag 2).

4.5.3 Subconclusie

De analyse van de 'observations' uit de drie cases naar benutting leverde de volgende constatering op. Ten eerste wordt duidelijk dat, hoewel er in de zin van het 'lessons learned' proces weliswaar geen sprake is van een 'geleerde les', er wel degelijk sprake is geweest van benutting. Een onderscheid tussen positieve en negatieve resultaten kon niet worden aangetoond. Bovendien werd vastgesteld dat het aantal positieve resultaten binnen de groep "observations" veel kleiner is dan het aantal negatieve resultaten. Voor wat betreft de benutting van minder complexe resultaten bleek benutting in alle gevallen instrumenteel van aard. Bij complexe resultaten was er in vier van de zes gevallen sprake van conceptuele benutting.

4.5.4 Lessons identified

In de geselecteerde cases zijn 29 resultaten geclassificeerd als 'lessons identified'. Deze resultaten zijn uitgesplitst naar positieve / negatieve resultaten en minder of meer complexe resultaten. Een weergave hiervan geeft onderstaande tabel:

Lessons Identified	
Positief/Minder complex	2
Positief/Meer complex	1
Negatief/Minder complex	11
Negatief/Meer complex	15
Totaal	29

4.5.5 Analyse

De resultaten die geclassificeerd zijn als 'lessons identified' hebben de analysefase (conform het 'lessons learned' proces) doorlopen. De achterliggende oorzaak van het probleem is bekend en er is sprake van een mogelijke oplossing. Daarnaast is duidelijk welke instantie (intern of extern de eigen organisatie), de oplossing moet implementeren (zie bijlage e, vraag 3).

Van de drie positieve resultaten is voor zowel de beide minder complexe resultaten als voor het meer complexe resultaat vastgesteld dat implementatie van de voorgestelde oplossing inmiddels was uitgevoerd. De oorzaak van het nog steeds te boek staan van deze resultaten als 'lessons identified' is te verklaren door een gebrek aan terugkoppeling en doordat controle op de uitvoering van de oplossing (en de controle of de oplossing daadwerkelijk het gewenste resultaat heeft opgeleverd) niet heeft plaatsgevonden. In termen van benutting zijn deze drie resultaten te kwalificeren als instrumenteel benut (zie bijlage e, vraag 3).

De 26 negatieve resultaten zijn uitgesplitst in 11 minder complexe en 15 meer complexe resultaten.

Voor wat betreft de minder complexe resultaten bleken zes resultaten in te gaan op het niet c.q. gebrekkig uitvoeren van bestaande regelgeving. In alle gevallen is door betrokken instantie beterschap beloofd. Controle op het juist toepassen van deze regelgeving heeft niet plaatsgevonden c.q. is een blijvende noodzaak, die zich buiten de 'lessons learned' organisatie afspeelt. De resultaten hebben geleid tot de notie dat bestaande regelgeving (met de attentie voor de specifiek in de resultaten genoemde zaken) moet worden toegepast. De resultaten lijken daarom het best te kwalificeren als instrumenteel benut. Vier resultaten betroffen de personele inrichting van het detachement. Interne afspraken c.q. het wijzigen van de omstandigheden hebben geleid tot het verhelpen of het verdwijnen van de relevantie van het (tijdelijk) probleem. Daadwerkelijke benutting van deze resultaten is niet aan te tonen. De laatste twee resultaten uit deze categorie betreft een

aanpassing in het kledingpakket voor uitgezonden militairen. Dit type aanpassing vergt veel tijd en speelt zich af buiten de eigen Luchtmachtorganisatie. Benutting van dit resultaat kon niet worden vastgesteld (zie bijlage e, vraag 3).

De 15 meer complexe resultaten bleken in de analyse terug te voeren naar een aantal thema's, waarbij steeds aspecten van een breder, onderliggend probleem werden genoemd. Ten eerste waren er vier resultaten die verband hielden met tekortkomingen in de logistiek (materieelbeheer). Vooral het goed afstemmen van verschillende bedrijfsvoeringssystemen – de Landmacht hanteert een ander systeem dan de Luchtmacht – bleek in deze gevallen deugdelijk beheer in de weg te staan. Het op elkaar afstemmen van beide systemen lijkt een gecompliceerde operatie, waarbij weliswaar de noodzaak is onderkend, maar dit heeft vooralsnog niet geleid tot structurele aanpassingen. Toch kan, met het oog op de ontstane discussie rondom dit thema, voor deze resultaten conceptuele benutting worden verondersteld. Een tweede thema betreft tekortkomingen in communicatie. Drie resultaten verwijzen er naar. De krijgsmacht gebruikt een grote verscheidenheid aan verbindingsmiddelen. In combinatie met de lastige omstandigheden in het inzetgebied heeft dit in enkele gevallen geleid tot minder dan optimale situaties. Oorzaken zijn divers. Het oplossen van deze problemen had inmiddels plaatsgevonden (instrumenteel). Het onderliggende probleem – ook in transitieperiodes dient goede communicatie te worden gegarandeerd – is door de resultaten scherp op het netvlies van betrokken afdeling gekomen. Voor wat betreft dit aspect is er sprake van conceptuele benutting. Thema drie betreffen operationele bemerkingen. De door deze bemerkingen gegenereerde discussies hebben niet geleid tot een aanpassing van het gehanteerde beleid. Er is daarom in deze (drie) gevallen geen sprake van instrumentele benutting. Toch kan worden betoogd dat de discussie (en het daaruit voortvloeiende besluit om het gehanteerde beleid niet aan te passen), te kenmerken is als conceptuele benutting. De overige vijf resultaten zijn divers van aard. Twee resultaten betroffen zaken waarvan een eventuele correctie (ver) buiten het bereik van de eigen organisatie ligt. Voor deze resultaten heeft benutting niet plaatsgevonden c.q. kan benutting niet plaatsvinden. Eén resultaat had dezelfde inhoud als een eerder genoemd resultaat (ATF-15/011) en had uit de database moeten worden verwijderd. Twee resultaten hadden hun relevantie verloren door gewijzigde omstandigheden (zie bijlage e, vraag 3).

4.5.6 Subconclusie

De analyse van de 'lessons identified' maakt duidelijk dat er een groot verschil bestaat tussen de stap van 'lessons identified' naar 'lessons learned' en de vraag of er sprake is geweest van benutting. In de meeste gevallen is vastgesteld dat er wel degelijk sprake is geweest van benutting. Voor een deel is dit te verklaren door administratieve oorzaken, maar zeker ook door het definitieverschil tussen 'geleerde les' en 'benutting'. Het onderscheid tussen positieve en negatieve resultaten kon niet worden aangetoond. Wel kon worden vastgesteld dat het aantal positieve resultaten binnen de groep 'lessons identified' veel kleiner is dan het aantal negatieve resultaten. Voor wat betreft de benutting van minder complexe resultaten bleek – indien sprake was van benutting – deze instrumenteel van aard. Voor wat betreft de complexe resultaten kon worden vastgesteld dat er naast instrumentele benutting in een aantal gevallen ook sprake was van conceptuele benutting.

4.6 Conclusie

Ten einde een goede fundering voor het onderzoek te creëren, zijn de database en de geselecteerde cases nader onderzocht. De database is voor wat betreft de cases 'opgeschoond' door de resultaten van de helikoptereenheid, die was toegevoegd aan ATF-15, te verwijderen. Daarnaast zijn de categorieën 'incident' en 'afgedaan' verwijderd. Vervolgens is geconcludeerd dat de onderzochte

cases (ondanks grote onderlinge verschillen) een soortgelijke verdeling volgen als de gehele database en dat de gekozen cases representatief zijn voor het geheel. Daarnaast is geconstateerd dat de gekozen cases qua timing enerzijds voldoende tijd bieden om het proces te doorlopen en anderzijds voldoende actueel zijn, waarmee 'verlies van geheugen' is voorkomen.

Na indeling naar positieve / negatieve resultaten en minder / meer complexe resultaten lijkt er een significant verschil te bestaan tussen positieve en negatieve resultaten die leiden tot 'geleerde lessen'. Positieve resultaten leiden in meer dan 50% van de gevallen tot 'best practices'. Bij negatieve resultaten is dit percentage aanmerkelijk lager. Daarnaast kan worden geconcludeerd dat minder complexe resultaten uit de onderzochte cases eerder leiden tot geleerde en geborgde lessen.

Voorts is vastgesteld dat alle 'geleerde lessen' (14) instrumenteel van aard zijn. Slechts in één geval was er ook sprake is andere vormen van benutting. De aard van de resultaten lijkt in deze 14 gevallen geen factor van invloed.

De analyse van de resultaten, die niet hebben geleid tot een 'geleerde les', leverde een geheel ander beeld op. Uit de analyse van zowel de 'observations' als de 'lessons identified' kon worden vastgesteld dat, hoewel in de zin van het 'lessons learned' proces weliswaar geen sprake is van een 'geleerde les', er wel degelijk sprake was van benutting. Daarnaast kon worden vastgesteld dat het aantal positieve resultaten binnen beide categorieën veel kleiner was dan het aantal negatieve resultaten. Een onderscheid in termen van benutting tussen positieve en negatieve resultaten kon in het onderzoek niet worden aangetoond. Voor wat betreft de benutting van minder complexe resultaten bleek deze benutting – indien sprake was van benutting – instrumenteel van aard. Voor wat betreft de complexe resultaten kon worden vastgesteld dat er naast instrumentele benutting in een aantal gevallen ook sprake was van conceptuele benutting en in één geval kon symbolische benutting worden vastgesteld. Een totaalbeeld van de 56 onderzochte resultaten is ondergebracht in onderstaande tabel:

Resultaten	Instrumenteel	Conceptueel	Symbolisch	Geen	Aantal
Positief/Minder complex	8+3+2				13
Positief/Meer complex	0+0+1				1
Negatief/Minder complex	4+4+6			5	19
Negatief/Meer complex	2+1+3	1*+4+7+3*	1*	1+5	23
Totaal					56

* betreffen resultaten met meerdere vormen van benutting en zijn maar 1 maal meegeteld.

Ten aanzien van de complexiteit kon worden geconcludeerd dat in het onderzoek minder complexe resultaten uitsluitend hebben geleid tot instrumentele benutting. Voor meer complexe resultaten was er ook sprake van conceptuele benutting en in één geval van symbolische benutting.

Op grond van de resultaten kon tevens worden geconcludeerd dat positieve resultaten in dit onderzoek instrumenteel zijn benut. Bij negatieve resultaten worden meerdere vormen van benutting gevonden. Hoewel er in de meeste gevallen sprake is van instrumentele benutting, kon in een aantal gevallen conceptuele benutting worden vastgesteld en bij één resultaat was er sprake van symbolische benutting. Daarnaast bleek voor vijf resultaten dat er per resultaat sprake was meerdere vormen van benutting.

5 Conclusies, discussie en aanbevelingen

5.1 Inleiding

Als onderdeel van de overheidsorganisatie beschikt ook de krijgsmacht over vele vormen van controle op basis van evaluaties. Naast verantwoording van gevoerd beleid en effectiviteitsmetingen, kent de krijgsmacht specifieke evaluaties die inzicht moeten geven in de operationele slagkracht van militaire eenheden. De afgelopen jaren is er – mede ingegeven door internationale ontwikkelingen op dit terrein – steeds meer aandacht voor het ‘leren’ van de ervaringen bij operationele inzet van de krijgsmacht.

Nederlandse luchtmachteenheden zijn de afgelopen twintig jaar regelmatig ingezet als bijdrage van Nederland aan VN- of NAVO vredesoperaties. Sinds 2010 dienen deze eenheden hun ervaringen vast te leggen conform de ‘lessons learned’ systematiek. Doel is steeds van de ervaringen te leren en zo nodig deze ervaringen te vertalen naar verbeteringen in werkwijze, organisatie of optreden.

Het gebruik van ervaringen (‘use’) wordt in de wetenschappelijke literatuur als één van de belangrijkste doelen van evalueren gezien. Voor dit onderzoek is er daarom gekozen om de factor ‘findings’ uit de categorie ‘Evaluation implementation’ (raamwerk Cousins en Leithwood, 1986) te beschouwen. Er is – mede met het oog op de doelstelling van het onderzoek – gekozen te bezien of de aard van die resultaten invloed heeft op de wijze waarop de resultaten worden benut.

Ten aanzien van de aard van de resultaten is ten eerste gekeken naar een mogelijk verschil tussen positieve (‘best practices’) en negatieve resultaten (verbeterpunten). Daarnaast is bezien in hoeverre de implementatiecomplexiteit van invloed is op de benutting.

Voor wat betreft de wijze van benutting richt dit onderzoek zich op de drie ‘klassieke’ vormen van benutting. De instrumentele, conceptuele en symbolische (of politieke) benutting van de resultaten. Andere vormen van benutting liggen minder voor de hand indien gekeken wordt naar het type van evaluatie waartoe de ‘lessons learned’ systematiek behoort. Daarnaast zijn andere vormen van benutting in veel gevallen slechts variaties op de genoemde ‘klassieke’ drie.

Met dit onderzoek is antwoord gegeven op de centrale onderzoeksvraag:

**In hoeverre bepaalt de aard van de resultaten die voortkomen
uit het ‘lessons learned’ proces de wijze waarop benutting plaatsvindt?**

5.2 Conclusies

Ten aanzien van de implementatiecomplexiteit bleek uit het onderzoek dat minder complexe resultaten uitsluitend hebben geleid tot instrumentele benutting. Voor meer complexe resultaten is in de meeste gevallen weliswaar sprake van instrumentele benutting, maar in een aantal gevallen kon ook conceptuele benutting worden vastgesteld en bij één resultaat was er sprake van symbolische benutting. Daarnaast waren er een aantal resultaten, waarbij er per resultaat meerdere vormen van benutting konden worden vastgesteld.

Het onderzoek heeft daarnaast aangetoond dat alle positieve resultaten uit de cases instrumenteel worden benut. Bij de negatieve resultaten zijn ook andere vormen van benutting gevonden. Er was bij een aantal resultaten sprake van conceptuele benutting en in één geval van symbolische benutting.

Daarnaast kon worden vastgesteld dat positieve resultaten vaker leiden tot benutting dan negatieve resultaten. Ten aanzien van resultaten met een meer of mindere implementatiecomplexiteit werd geen significant verschil gevonden.

Tot slot is tijdens het onderzoek gebleken dat er bij veel resultaten uit het 'lessons learned' proces weliswaar geen sprake is van een 'geleerde les', er wel degelijk benutting plaatsvindt.

5.3 Verwachtingen

In de inleiding van deze studie zijn verwachtingen uitgesproken over mogelijke uitkomsten van het onderzoek. Hier is betoogd dat eenvoudig te implementeren resultaten vermoedelijk instrumenteel van aard zullen zijn. De gevonden uitkomsten van het onderzoek komen in grote lijnen overeen met deze verwachting. Onverwacht – maar goed verklaarbaar – was de uitkomst dat positieve resultaten vaker leiden tot benutting dan negatieve resultaten. Het vermoeden bleek juist dat, hoewel een resultaat niet had geleid tot een 'geleerde les', er toch met een hoge mate van waarschijnlijkheid benutting kon worden aangetoond.

5.4 Relevantie

Ten aanzien van de wetenschappelijke relevantie van het onderzoek, kan worden opgemerkt dat het onderzoek heeft bijgedragen aan de door Christie (2007) genoemde smalle empirische basis. De conclusies van het onderzoek komen voor wat betreft de wijze waarop benutting plaatsvindt, overeen met de gevonden resultaten uit eerder onderzoek Preskill zoals genoemd in het artikel van Bober en Bartlett (2004).

Kijkend naar de maatschappelijke relevantie lijkt het onderzoek aan te tonen dat positieve resultaten ('best practices') eerder en sneller leiden tot concrete verbeteringen in werkwijze, processen en/of procedures. Daarnaast lijkt een beoordeling van het 'lessons learned' proces als geheel op grond van uitsluitend de 'geleerde lessen', het proces tekort te doen. Die conclusie zou eenvoudig kunnen luiden dat slechts een beperkt deel van de resultaten leidt tot gebruik en dat het vaak kleine, eenvoudig te implementeren resultaten betreffen. Indien echter wordt gekeken naar de benutting van deze resultaten kan in veel gevallen worden aangetoond dat benutting wel degelijk plaatsvindt en ook meer complexe resultaten leiden tot enige vorm van benutting. Een nadere uitwerking van deze constatering zal plaatsvinden in de paragraaf 'discussie'.

5.5 De onderzoeksresultaten ten opzichte van het theoretisch kader

'Use' of 'utilization' vormen de kern van het huidige denken over evaluaties. De gedachte dat evaluaties nuttig dienen te zijn staat centraal. Bober en Bartlett (2004) stellen dat – wil een evaluatie echt een formatief of summatief doel dienen – de resultaten moeten worden benut. Relevant voor het onderzoek is hun verwijzing in het artikel naar eerder onderzoek dat is uitgevoerd door Preskill in 1984 (Preskill, 1984). Zij constateert dat evaluatieresultaten zowel van instrumentele als van conceptuele aard kunnen zijn. Echter, een veel groter percentage van benutting bleek van instrumentele aard te zijn. Preskill realiseerde zich dat het ontdekken van benutting van conceptuele

aard veel lastiger is. Ze was er echter van overtuigd dat, indien ze betere toegang tot het hogere management had gehad, ze deze vorm van benutting had kunnen identificeren en bestuderen. Weiss, Murphy-Graham en Birkeland (2005) voerden onderzoek uit naar het D.A.R.E.-programma. In deze studie werd aangetoond dat resultaten van een evaluatie (de 'findings') een dominante factor zijn in de benutting van de evaluatie. Relevant voor het onderzoek zijn ook de opmerkingen die Weiss e.a. maken ten aanzien van het steeds meer in zwang zijnde 'best practices' model. Hierbij worden programma's onderzocht met als doel de meest effectieve/efficiënte werkwijzen te identificeren. Weiss e.a. stellen dat deze vorm van evaluatie uitnodigt om druk uit te oefenen om de 'best practices identified' breed toe te passen (Weiss e.a 2005. p. 26).

De resultaten van het uitgevoerde onderzoek bevestigen de uitkomsten uit het onderzoek van Preskill. Ten eerste kon tijdens het onderzoek worden vastgesteld dat er bij een hoog percentage resultaten sprake was van instrumentele benutting. Ten tweede, dat er ook andere vormen van benutting konden worden geïdentificeerd. Tot slot is in het onderzoek duidelijk geworden dat er terecht aandacht bestaat voor positieve ervaringen ('best practices').

5.6 Betrouwbaarheid en validiteit van onderzoek

Het onderzoek is uitgevoerd als meervoudige gevalsstudie, waarbij gelijkvormige cases in de periode van eind 2010 tot eind 2011 zijn onderzocht. Door het gelijkvormig maken van deze cases (het betreffen 3 opeenvolgende F-16 detachementen) is het mogelijk geweest bij de analyse gebruik te maken van herhalingslogica. Met deze keuze is tevens een balans gevonden waarbij enerzijds voldoende tijd is gegeven om te bezien op welke wijze benutting heeft plaatsgevonden, anderzijds zijn de cases voldoende actueel, waarmee 'verlies van geheugen' is voorkomen.

Voor het onderzoek zijn een drietal onderzoekbronnen gebruikt. Er is gebruik gemaakt van de 'lessons learned' database. Deze database bevat alle resultaten uit de evaluaties. Daarnaast is in de database opgenomen of, en zo ja in hoeverre, de resultaten hebben geleid tot verandering c.q. zijn geïmplementeerd. De 'lessons learned' database is omgevormd tot – en gebruikt als – een 'case study' database. De tweede bron van informatie werd gevormd door de evaluatieverslagen zelf. Deze verslagen (en ander achtergrond materiaal) zijn gebruikt om een diepere analyse uit te voeren ten aanzien van de gevonden resultaten. De derde bron van informatie werd gevormd door de gevoerde gesprekken. Deze gesprekken hebben bijgedragen aan een breder en scherper inzicht en hebben gefungeerd als toetsingsmechanisme voor de geanalyseerde data. Met de gekozen opzet is voldaan aan de drie principes ten aanzien van dataverzameling en is de betrouwbaarheid van het onderzoek verhoogd. Het gebruik van meerdere bronnen, heeft daarnaast triangulatie mogelijk gemaakt en heeft bijgedragen aan het creëren van een 'chain of evidence'.

Ten aanzien van de validiteit van het onderzoek is – zoals hierboven beschreven – gebruik gemaakt van verschillende bronnen. Er is getracht van de conclusies en uitspraken een consistent en samenhangend geheel te maken. Hierbij is gebruik gemaakt van pattern matching en explanation building. Daarnaast zijn strijdige verklaringen getoetst en nader geanalyseerd door deze tijdens de gesprekken voor te leggen. De gekozen cases worden representatief voor de gehele database geacht. Dit is niet alleen kwantitatief zichtbaar (zie tabel para 4.2.) maar wordt ook bevestigd door 'lessons learned' functionaris (zie bijlage c). Hierdoor lijkt het dat aan de resultaten van het onderzoek een bredere reikwijdte kan worden toegekend.

5.7 Beperkingen van het onderzoek

Hoewel er veel aan gedaan is om een zo valide en betrouwbaar onderzoek uit te voeren, kent het onderzoek een aantal beperkingen. Zonder volledig uitputtend te kunnen zijn, worden in deze paragraaf een paar in het oog springende beperkingen genoemd. Ten eerste de gekozen cases. Hoewel er met zorg is gekozen en deze keuze representatief lijkt voor het geheel, is dit niet met zekerheid vast te stellen. Daarvoor zouden alle resultaten uit de database moeten worden beschouwd. Ten tweede heeft een subjectieve beoordeling en indeling plaatsgevonden, vooral waar het gaat om de indeling meer of minder complex. Hierbij wordt tevens opgemerkt dat het onderzoek de aard van de resultaten slechts op twee dimensies heeft beschouwd. Ook de toekenning van het soort benutting, heeft plaatsgevonden op grond van eigen analyse en aan de hand van de gevoerde gesprekken. Deze vorm van informatieverzameling en oordeelsvorming heeft per definitie een subjectief karakter. Ten derde worden in het onderzoek uitspraken gedaan op grond van slechts een beperkt aantal resultaten. Generalisatie van deze uitspraken dient dan ook met de nodige voorzichtigheid en bescheidenheid plaats te vinden.

5.8 Discussie naar aanleiding van het onderzoek

5.8.1 Aandacht voor positieve resultaten

De aandacht voor positieve resultaten lijkt terecht. Het onderzoek lijkt de kracht van het innovatieve vermogen op de 'werkvloer' te bevestigen. Hierbij betreft het niet alleen het verbetervoorstel op zich. In veel gevallen heeft ook de implementatie van het voorstel al plaatsgevonden. Hierdoor resteert alleen een goede borging.

5.8.2 'Geleerde les' ten opzichte van 'benutting'

Als 'bijvangst' van het onderzoek is geconstateerd dat er een groot verschil bestaat tussen de mate waarin resultaten leiden tot 'geleerde lessen' en de mate waarin benutting van resultaten kan worden vastgesteld. Het 'lessons learned'- proces leidt slechts voor een beperkt deel van de resultaten (14 van 56) tot gebruik ('geleerde les'). Daarnaast betreffen deze 'lessen' vaak kleine, eenvoudig te implementeren resultaten. Wordt echter gekeken naar de benutting van deze resultaten levert dit een heel ander beeld op. Van de 56 resultaten kan benutting in 45 gevallen worden aangetoond en blijken ook meer complexe resultaten te leiden tot enige vorm van benutting. (H)erkenning van benutting blijkt slechts beperkt plaats te vinden.

5.8.3 Aandacht voor het programma

Er lijkt binnen overheidsorganisaties (en dus ook de Luchtmacht) blijvend aandacht voor het verbeteren van de eigen prestatie. Er kan echter ook worden geconstateerd dat er – als gevolg van de recente bezuinigingen – een tendens waarneembaar is, om de capaciteit nodig is om deze programma's uit te voeren te beperken. Het lijkt dus van belang het nuttig effect van evaluatieprogramma's helder en volledig binnen de organisatie te communiceren (zie paragraaf 2.3.). Dat deze constatering ook geldt voor het 'lessons learned' programma binnen de Luchtmacht mag blijken uit de antwoorden van de Hr. Op den Kamp (bijlage d) en de opmerkingen van de Hr. Van Hees (bijlage c, vraag 11).

5.9 Aanbevelingen naar aanleiding van het onderzoek

5.9.1 Praktische implicaties volgend uit het onderzoek

Op grond van de resultaten van dit onderzoek zou een ander beeld ten aanzien van het nut van dit type van evaluaties kunnen ontstaan. Eerst binnen de 'lessons learned' organisatie zelf vanuit het

besef dat er ook zonder ‘geleerde les’ sprake kan zijn van benutting. Daarnaast zou – gezien de kracht van positieve resultaten – de focus op ‘best practices’ verder kunnen worden geëntameerd. Goede, regelmatige communicatie met de diverse ‘stakeholders’ lijkt essentieel voor het welslagen van het proces per resultaat én de waardering voor het programma als geheel.

5.9.2 *Aanleiding voor vervolgonderzoek*

In het verlengde van de onderzoeksresultaten lijkt vervolgonderzoek vooral interessant op het terrein van het ‘lessons learned’ proces zelf. In dit onderzoek hebben de evaluatieresultaten in combinatie met de drie ‘klassieke’ vormen van benutting een centrale plaats ingenomen. Een beschouwing van de mate waarin/de wijze waarop dit proces op zich leidt tot benutting, zou een waardevolle aanvulling zijn op het beter begrijpen van dit type van evaluatie.

Jan van der Veer, 13 juni 2013

Literatuurlijst

Alkin, M.C. and Christie, C.A. (2004)

An Evaluation Theory Three

Alkin, M.C. and Taut, S.M. (2003)

Unbundling evaluation use

Studies in educational Evaluation 29, pp. 1 - 12

Balthasar, A. and S. Rieder (2000)

Learning from evaluations: effects of the evaluation of the Swiss Energy 2000 Program

Evaluation, 7 2000; vol. 6: pp. 245 - 260.

Baltasar, A. (2009)

A contribution to a theory of evaluation influence based on Swiss Experience

Evaluation Review 3, vol. 33: pp. 226 - 256

Bober, C. and Bartlett, K (2004)

The utilization of training program evaluation in corporate universities.

Human Recourse Development Quarterly, vol. 15 no. 4, winter 2004

Burke Johnson, R. (1998)

Toward a theoretical model of evaluation utilization

Evaluation and Program Planning Vol. 21, Issue 1, February 1998: pp. 93-110.

Christie, C. (2007)

Reported Influence of Evaluation Data on Decision Makers' Actions

An Empirical Examination

American Journal of Evaluation March 2007 vol. 28 no. 1: pp. 8-25

Cousins, J.B. and Leithwood K.A. (1986)

Current empirical research on evaluation utilization

Review of educational research, 56: pp. 331 - 364

Dal Santo, T, Goldberg, S, Choice, P and Austin, M.J. (2002)

Exploratory Research in Public Social Service Agencies: As Assessment of Dissemination and Utilization

Journal of Sociology and Social Welfare, December, 2002, Volume XXIX, Number 4

Foss, K., Rebien, C., Carlsson, J. (2002)

Process use of evaluations, types of use that precede lessons learned and feedback.

Evaluation, vol. 8: pp. 29-45.

Greene, J. (1988)

Stakeholder Participation and Utilization in Program Evaluation,

Evaluation Review, 4 1988; vol. 12: pp. 91 - 116.

- Henry, G. and M. Mark (2003a)
Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions,
American Journal of Evaluation, 9 2003; vol. 24: pp. 293 - 314.
- Henry, G. and M. Mark (2003b)
Toward an agenda for research of evaluation,
New directions for evaluation, 97: pp. 69 – 80.
- Henry, G. and Rog D.J. (1998)
A realist theory and analysis of utilization. In Realist evaluation: An emerging theory in support of practice,
ed. Gary T. Henry, George Julnes, and Melvin M. Mark, 89-102. San Francisco: Jossey-Bass.
- Janowitz, M. (1970)
Sociological models and social policy. In Political conflict: Essays in political sociology,
ed. Morris Janowitz, 243-59. Chicago: Quadrangle Books.
- Johnson, K., Greenseid, L., Toal, S., King, J., Lawrenz, F., Volkov, B. (2009)
Research on Evaluation Use: A Review of the Emperical Literature From 1986 to 2005
American Journal of Evaluation 2009 30: pp. 377
- Leviton, L. and E. Hughes (1981)
Research On the Utilization of Evaluations: A Review and Synthesis
Evaluation Review, 8 1981; vol. 5: pp. 525 - 548.
- Leviton, L. (2003)
Evaluation Use: Advances, Challenges and Applications
American Journal of Evaluation, 12 2003; vol. 24: pp. 525 - 535.
- Patton, M. (1997)
Utilization-focused evaluation: The new century text: pp. 301-338.
Thousand Oaks, CA: Sage.
- Patton, M. (1998)
Discovering process use.
Evaluation, vol. 4: pp. 225-233.
- Reineke, R. A. (1991)
Stakeholder involvement in evaluation: Suggestions for practice.
American Journal of Evaluation, 12: pp. 39-44.
- Rich, R.F. (1991)
Knowledge Creation, Diffusion, and Utilization Perspectives of the Founding Editor of Knowledge
Science Communication March 1991 vol. 12 no. 3 pp. 319-337
- Rossman, G. B., and Rallis, S. F. (2000)
Critical inquiry and use as action.
New Directions for Evaluation, 88: pp. 55-69.

- Striven, M. (2007)
Activist evaluation.
Journal of multidisciplinary evaluation, 4 (7): pp. i - ii
- Shulha, L. and J. Cousins (1997)
Evaluation Use: Theory, Research, and Practice Since 1986,
American Journal of Evaluation, 2 1997; vol. 18: pp. 195 - 208.
- Turnbull, B. (1999)
The mediating effect of participation efficacy on evaluation use.
Evaluation and Program Planning, 22: pp. 131-140.
- Weiss, Carol H. (1977)
Research for policy's sake: The enlightenment function of social research.
Policy Analysis 3: 531-45.
- Weiss, C. (1998)
Have We Learned Anything New About the Use of Evaluation?
American Journal of Evaluation, 3 1998; vol. 19: pp. 21 - 33.
- Weiss, Murphy-Graham and Birkeland (2005)
An alternate route to policy influence
American Journal of Evaluation, Vol. 26 No.1, March 2005: pp. 12-30
- Werkman R.A., J.J. Boonstra en K.M. Bennebroek Gravenhorst (2001)
Het verandervermogen van organisaties
M&O nummer 2 – maart/april - 2001
- Yin, R.K., (2009)
Case Study Research, Design and Methods.
Thousand Oaks: SAGE, 4th edition.

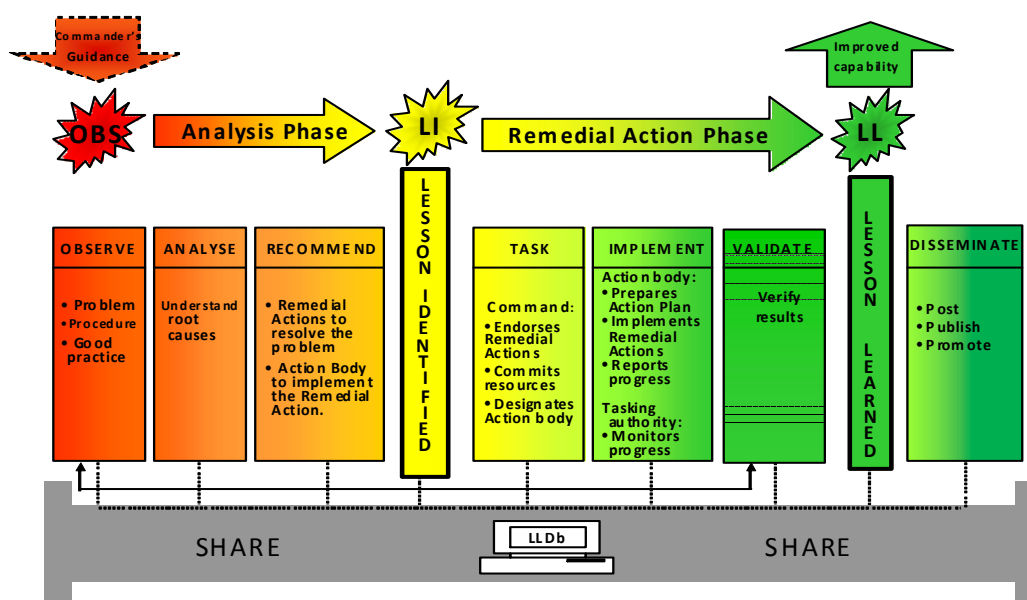
Bijlage A: Explicatie 'lessons learned' systematiek

Inleiding

Deze bijlage gaat in op de systematiek van het 'lessons learned' proces, zoals die wordt toegepast binnen de krijgsmacht en dus ook binnen de Luchtmacht.

Het proces

Het proces dat wordt gehanteerd is een vertaling van het door, de NAVO gestandaardiseerde 'lessons learned' proces. De gedachte achter het 'lessons learned' proces is dat in een organisatie, door een gestructureerde manier van leren, zowel individuen als de organisatie als geheel fouten uit het verleden kunnen vermijden. Daarnaast kan de toepassing van 'best practices' bijdragen tot succes. In een militaire context betekent dit niet alleen dat operationele risico's kunnen worden verkleind, maar dat ook de operationele effectiviteit wordt verhoogd.



Toelichting op bovenstaand schema - fasering

Het doen van een 'observation'

Een 'observation' is een beschrijving van een activiteit, voorval of werkwijze, die wordt gedocumenteerd met relevante informatie over de factoren die van invloed zijn geweest. Het kan zijn een gebeurtenis waarbij de uitkomst niet voldeed aan de verwachting (een zogenaamde 'bad practice'), maar ook iets wat in de praktijk succesvol bleek te zijn (een zogenaamde 'best practice'). In principe kan door iedereen een 'observation' worden gedaan. Om echter een *overload* aan 'observations' te voorkomen verdient het de aanbeveling om structuur aan te brengen in de informatiestroom.

Raw observations

'Observations' worden meestal aangeleverd in een incomplete vorm (ontbreken van de relevante factoren van invloed). In de 'lessons learned' terminologie zijn dit de *raw observations*. 'Observations' dienen door (of onder

begeleiding van) een 'lessons learned' functionaris verder te worden ontwikkeld tot *mature observations*, zo nodig ondersteund door een ter zake deskundige. De *mature observation* maakt de kern van het probleem duidelijk en geeft voldoende informatie en inzicht om oorzaak en gevolg van datgene wat geconstateerd is te begrijpen.

De analysefase

In de analysefase wordt de dieperliggende oorzaak van het probleem of juist de '*best practice*' geïdentificeerd. Bij voorkeur wordt de analyse zo dicht mogelijk bij de bron van de 'observation' uitgevoerd. Als dit niet mogelijk is vanwege de complexiteit van het geconstateerde probleem / best practice, dient een verzoek naar een hoger niveau te worden gedaan voor aanvullende expertise of wordt de analyse door een hoger niveau overgenomen. Immers, de 'observation' is doorgaans slechts een constatering van een symptoom van een probleem of '*best practice*'. Bij de analysefase zal de 'lessons learned' functionaris wederom zo nodig specialistische kennis inschakelen om de achterliggende oorzaak te kunnen achterhalen. In deze fase wordt ook de mogelijke oplossing uitgewerkt en wordt geïdentificeerd welke instantie de voorgestelde oplossing zou moeten implementeren. De resultaten (Lessons Identified) van de analysefase zijn dan ook driedelig:

- a. De dieperliggende oorzaak van het probleem (Root Cause).
- b. Een voorgestelde oplossing (Remedial Action).
- c. De instantie die de oplossing zou moeten implementeren (Action Body).

Besluitvorming

Lessons identified worden aangeboden aan diegene (het management), die de middelen en mogelijkheden heeft om de voorgestelde oplossing te (laten) implementeren. Implementeren vraagt immers tijd, geld, personele en materiële middelen en (misschien wel de belangrijkste succesfactor) aandacht van het management.

Het implementatietraject.

Indien het management besluit om een Lesson Identified te implementeren (dat wil zeggen de Lesson Identified om te laten zetten tot een Lesson Learned), moet er een opdracht worden gegeven om de *Remedial Action* uit te voeren. Het verdient doorgaans aanbeveling om hierbij een projectmatige aanpak te hanteren. De 'lessons learned' functionaris kan daarbij de voortgang bewaken, maar zal niet optreden als projectmanager voor de verandering (*Lessons Learned is about managing the change for the commander*). Eventueel kan aan het eind van deze fase worden gevalideerd of de gekozen oplossing ook daadwerkelijk het probleem heeft opgelost.

Nadat de implementatie is voltooid, is de les geleerd. De geleerde les dient nu onder alle belanghebbenden te worden verspreid.

Bijlage B: Tabellen onderzochte resultaten

	Resultaat ATF 15	positief	negatief	Meer complex	Minder complex	Classificatie
1	002		X		X	LI
2	003		X		X	LI
3	004		X		X	LI
4	005		X	X		LL
5	006		X		X	LI
6	008		X		X	LI
7	009		X		X	LI
8	010		X		X	OBSERVATION
9	011		X		X	LL
10	014		X		X	LL
11	016	X			X	LI
12	017		X		X	LL
13	018		X	X		LI
14	019		X	X		LI
15	020		X	X		LI
16	024		X	X		LI

	Resultaat ATF 16	positief	negatief	Meer complex	Minder complex	Classificatie
1	002	X			X	BEST PRACTICE
2	003		X		X	LI
3	004	X			X	BEST PRACTICE
4	005	X			X	LI
5	006		X		X	OBSERVATIE
6	008		X		X	LI
7	009		X	X		OBSERVATION
8	010		X		X	LI
9	012	X		X		LI
10	013		X	X		LI
11	014		X	X		LI
12	016		X	X		LI
13	017		X	X		LI
14	018	X			X	BEST PRACTICE
15	019	X			X	BEST PRACTICE
16	020		X		X	OBSERVATION
17	021		X	X		OBSERVATION
18	022	X			X	BEST PRACTICE
19	023	X			X	BEST PRACTICE
20	024		X		X	LI
21	025	X			X	BEST PRACTICE
22	026		X	X		OBSERVATION
23	027	X			X	OBSERVATION
24	029	X			X	OBSERVATION
25	030	X			X	OBSERVATION
26	031	X			X	BEST PRACTICE
27	032		X	X		OBSERVATION
28	034		X	X		LI

	Resultaat ATF 17	positief	negatief	Meer complex	Minder complex	Classificatie
1	001		X	X		LI
2	002		X	X		OBSERVATION
3	004		X	X		LL
4	005		X	X		LI
5	006		X		X	OBSERVATION
6	007		X	X		OBSERVATION
7	008		X	X		LI
8	009		X	X		LI
9	010		X	X		LI
10	011		X		X	LI
11	012		X		X	LI
12	013		X	X		LI

Bijlage C: Gesprekverslag Hr. A.P. van Hees

(diverse gesprekken in de periode november – december 2012)

De Heer Van Hees is medewerker van het bureau Evaluaties en specialist op het gebied van de gekozen cases. Primair doel van de gesprekken is het verkrijgen van een goed inzicht in het 'lessons learned' proces en de beoordeling van de resultaten vanuit de 'lessons learned' systematiek. De heer Van Hees heeft een waardevolle bijdrage geleverd bij de analyse van elk van de 56 resultaten.

2. Kun je me uitleggen hoe de database is opgebouwd?

Antw: Ieder detachement is verplicht na afloop van de uitzendperiode een verslag te schrijven. Dit verslag lopen we na en halen er de tekortkomingen en goede zaken uit. Deze worden opgenomen als 'observation' in de database. Vervolgens sturen we het verslag, samen met de lijst van bemerkingen naar de verschillende afdelingen.

3. Hoe houden jullie de database bij?

Antw: De afdelingen geven een reactie op de bemerkingen. Ook deze verwerken we in de database. Vervolgens volgen we de vervolgacties en passen we de database aan.

4. Volgen jullie precies het proces zoals beschreven in het handboek? Zo nee, waaruit bestaan de afwijkingen?

Antw: Dat proberen we zoveel als dat mogelijk is. Soms kunnen we echter stapjes overslaan indien het erg duidelijk is dat we het hele proces niet hoeven te doorlopen. Bijvoorbeeld wordt bij eenvoudige 'best practices' de oplossing al meegestuurd en is ook duidelijk dat de verbetering al wordt gebruikt.

5. Ik heb voor het onderzoek drie cases geselecteerd. Wat vind je van deze keuze?

Antw: Ik denk een goede keuze. De drie detachementen hebben alle drie in Kandahar gezeten en ongeveer dezelfde taak uitgevoerd. Daarbij is het lang genoeg geleden om actie op de opmerkingen te mogen verwachten en ook weer niet zolang geleden dat we het allemaal weer zijn vergeten.

6. Ik heb de drie gelijkvormig gemaakt door de resultaten van ATF 15 te ontdoen van de helikopter eenheid. Heb ik dat correct gedaan?

Antw: Ja. Ten eerste waren opmerkingen van de helikopter eenheid als zodanig gemerkt. Daarbij komt dat ze op een andere locatie zaten en daarom toch andere zaken noemden.

7. Wat is de reden dat in de database de termen 'incident' en 'afgedaan' voorkomen?

Antw: Simpel. Het is voor ons een indicatie dat we er niet meer naar hoeven te kijken. Vaak omdat het een heel klein iets is of dat de omstandigheden zijn veranderd, of sowieso een eenmalig ding. We gebruiken deze termen om vervuiling in de database te voorkomen.

8. Welke definitie worden door jullie gehanteerd voor de begrippen 'incident' en 'afgedaan'?

Antw: Incidenten zijn eenmalige resultaten zonder verbeteractie, maar die we nog niet willen weggooien, voor het geval dat een volgend detachement het ook noemt en we het opnieuw moeten beoordelen. Afgedaan gebruiken we voor zaken die niet zullen leiden tot aanpassingen (eenmalig).

9. Voor het onderzoek heb ik deze categorieën uit de database verwijderd en resultaten zo nodig anders geclassificeerd. Is dit correct gebeurd?

Antw: Volgens mij is dit correct gebeurd.

10. Wat is je oordeel over de wijze waarop ik de database heb 'opgeschoond'?

Antw: Het resultaat lijkt me goed* en een goede basis voor je onderzoek (* de heer van Hees heeft ieder resultaat met mij samen beoordeeld).

11. Zijn de grote verschillen tussen de onderlinge cases te verklaren? Zo ja, welke verklaring?

Antw: Iedere commandant heeft de vrijheid om in het verslag die zaken te noemen die hij belangrijk vindt. Daarbij komt dat de ene commandant het verslag zelf schrijft en de andere commandant zijn staf erbij betreft. Vaak bepaalt ook de achtergrond van de commandant de inhoud van het verslag. De ene heeft een puur operationele achtergrond en de andere heeft ook op de staf gezeten (waardoor hij vaak op de hoogte is van de achtergrond van een bepaald probleem. Zelfs de grondhouding van de commandant speelt een rol: de een kijkt sneller naar positieve zaken, terwijl de ander meer aandacht heeft voor tekortkomingen. Zelfs de aandacht voor en de tijdigheid van aanleveren van het verslag speelt een rol.

12. Van de 56 resultaten zijn 13 resultaten nog steeds gemarkeerd als 'observations'. Hoe verklaar je dat (met het oog op de gekozen periode)? Wat is hiervan de oorzaak?

Antw (schriftelijk): Verdere analyse van deze 13 'observations' heeft het volgende opgeleverd: een aantal (5) liggen boven het CLSK niveau, andere (3) zijn in een later stadium aangemerkt als 'lessons identified', 3 zijn te beschouwen als incident, door verandering van missie, procedure (anders dan door de 'observation') en 2 hebben minimale impact op de organisatie dat dit 'de moeite niet loont'. Waarom nog steeds gelabeld als 'observation'? a) door meer dan halvering van 'lessons learned' bureau; b) doordat een nagenoeg zelfde 'observation' in een latere missie wel is verwerkt tot 'lessons identified' (en daarbij de eerdere niet is meegenomen). De individuele 'observations' zijn in deze database niet aan elkaar te koppelen, waardoor dit voorkomen kan worden.

13. Van de 56 resultaten zijn 29 gemarkeerd als 'lessons identified'. Waarom zijn deze resultaten (nog) niet omgezet naar 'lessons learned'/'best practices'? Wat is hiervan de oorzaak?

Antw (schriftelijk): Velerlei redenen liggen hieraan ten grondslag. Hoofdoorzaken zijn echter:

- a) er vindt nauwelijks terug koppeling plaats over de daadwerkelijke borging;
- b) veel 'observations' worden ook parallel aan het 'lessons learned' traject doorgegeven aan actienemers en 'verdwijnen uit het zicht';

- c) 'lessons learned' neemt (nog) geen centrale plaats in binnen de organisatie;
- d) door de reductie van het 'lessons learned' bureau wordt het bijna onmogelijk om elke 'lesson identified' te 'tracken & traceren'.

14. Daarnaast kan worden geconcludeerd, dat van het geheel aan 'observations' (56), (slechts?) 14 resultaten hebben geleid tot een 'geleerde les' (6 'lessons learned' en 8 'best practices'). Hoe is dit te verklaren?

Antw (schriftelijk): Uit ervaringsgetallen binnen NAVO blijkt dat gemiddeld minder dan 25% resulteert in een 'lessons learned'/'best practices'. Verder blijft de organisatie zich aanpassen aan de situatie / verbeteren. Een 'lessons identified' kan dus best al een 'lessons learned' zijn geworden. Verder zijn een aantal 'observations' uiteindelijk van een weinig belang dat verdere acties niet genomen worden. Ook wordt soms vanwege procedurele gelijkheid tussen de OPCO's (noot: Operationeel Commando) of een procedure van de CLAS (noot: het Commando Landstrijdkrachten) als het coördinerend OPCO niet aan elk voorstel uitvoering gegeven. Verder, op het einde van de missie, worden 'observations' die tijdens de missie zijn opgebracht en ook reeds opgelost, veelal niet meer opgebracht. Dit geldt vooral voor tactisch operationele, geneeskundige en (bedrijfs) veiligheid 'observations'.

15. Wat is je oordeel over de criteria die ik heb gebruikt om de implementatie complexiteit van een resultaat in te delen?

Antw: Naast de twee door jou geformuleerde criteria, is er nog een derde van belang. Vaak moet er rekening worden gehouden met regelingen, procedures of beleid die wordt toegepast in andere delen van de organisatie. Het uitgangspunt dat regelingen voor de gehele organisatie hetzelfde zijn, kan betrekkelijk eenvoudige zaken toch ingewikkeld maken om te realiseren.

16. Heb ik correct ingedeeld?

Antw: Natuurlijk zit er een subjectief element in een dergelijk indeling. Toch lijkt me de indeling goed verdedigbaar.

17. Hoe is te verklaren dat in geval van positieve resultaten er een veel groter deel daadwerkelijk leidt tot een 'best practice'?

Antw (schriftelijk): Vaak wordt een positieve observatie met een kleine omvang en vergezeld van een voorstel hoe dit door te voeren en te borgen, direct al ingevoerd als 'best practice' zonder de volledige 'lessons learned' cyclus te doorlopen. In veel gevallen is ten tijde van de rapportering de 'observation' al doorgevoerd in de organisatie.

18. Op welke wijze worden positieve resultaten benut? Zo ja hoe? Zo nee, waarom niet?

Antw: Positieve resultaten worden zo mogelijk doorgevoerd of als aanbeveling meegegeven aan het volgende detachement. Deze 'best practices' zijn eenvoudig van uitvoer en lokaal en dus op kleine schaal te implementeren. Dit komt de snelheid ten goede.

19. Op welke wijze worden negatieve resultaten benut? Zo ja hoe? Zo nee, waarom niet?

Antw: In dit geval is de voorgestelde aanbeveling, niet overgenomen omdat blijkbaar de rapporteur niet het volledige overzicht had. Wel heeft dit geleid tot een verduidelijking van de reden waarom voor een oplossing gekozen is.

20. Op welke wijze worden minder complexe resultaten benut? Zo ja hoe? Zo nee, waarom niet?

Antw: Zijn eenvoudig door te voeren en vaak op lokale of beperkte schaal.

21. Op welke wijze worden complexe resultaten benut? Zo ja hoe? Zo nee, waarom niet?

Antw: Complexe zaken worden ten dele benut. Vaak liggen deze boven het CLSK niveau (noot: Commando Luchtstrijdkrachten) en vindt er een duidelijke terugkoppeling plaats waarom een aanbeveling wel of niet wordt overgenomen. Vaak heeft het doorvoeren van een aanbeveling andere negatieve implicaties die de rapporteur niet overziet.

Tijdens de gevoerde gesprekken is ieder resultaat separaat besproken. Hierdoor is niet alleen de indeling van de resultaten – zoals gebruikt in het onderzoek – tot stand gekomen, maar is ook een eerste inschatting gemaakt ten aanzien van het type benutting. Weergave van deze analyse is niet opgenomen in dit verslag. Het betroffen in de meeste gevallen specifiek interne zaken (met de daarbij behorende terminologie) en in sommige gevallen met een vertrouwelijk karakter.

Akkoord met 1 aanvulling (vraag 11): 27 maart 2013

Bijlage D: Gesprekverslag Hr. E.F. op den Kamp

(diverse gesprekken in de periode januari 2013)

Weergave van de gesprekken met de heer Op den Kamp, Hoofd bureau Evaluaties, waaronder ook de 'lessons learned' evaluaties worden uitgevoerd. De met hem gevoerde gesprekken betroffen voornamelijk bespiegelingen over het 'lessons learned'- proces als geheel. De gestelde vragen beoogden inzicht te bieden in de wijze waarop het evaluatieprogramma binnen de organisatie wordt gewaardeerd. Naast de gesprekken zijn een aantal vervolgvragen door hem schriftelijk beantwoord.

1. Het merken van een resultaat als 'best practice' of 'lesson learned' impliceert dat het resultaat heeft geleid tot een daadwerkelijke verbetering. Is dit zo? Hoe wordt dit vastgesteld?

Antw: Nadat de vigerende regelgeving is aangepast. Verder wanneer een bepaalde 'observation' niet terugkomt zou een mogelijke reden kunnen zijn dat de verandering het gewenste resultaat geeft.

2. Worden 'geleerde lessen' voldoende geborgd? En zo ja, op welke wijze? Zo nee, wat is de oorzaak?

Antw: Het bureau 'lessons learned' heeft geen volledig zicht op alle 'lessons learned' en hun borging. Dit omdat de terugkoppeling niet (altijd) gebeurt. Borging zien we bv. in veranderingen in de Operationele Aanwijzing, de organisatietabel of andere vigerende regelgeving.

3. Hoe groot wordt het belang van evaluaties binnen de organisatie geacht?

Antw: Het belang dat de organisatie hecht aan 'lessons learned' is groot. Zo zijn er diverse afzonderlijke 'lessons learned' trajecten die inmiddels hun voetsporen hebben verdiend (TACTESS, VKAM, BRW). Echter de hier bedoelde 'lessons learned' organisatie is nog redelijk nieuw. Het belang hiervan wordt wel steeds groter omdat in een beginfase de 'acceptatie' nog moet komen om iets te evalueren of om 'ondervraagd' te worden over de reden van een procedure of situatie. Na de acceptatiefase zien we nu een participatie in het 'lessons learned' traject.

4. Uit dit belang zich in voldoende aandacht voor het programma?

Antw: Hierin enkel het bureau 'lessons learned' bekijkend is de aandacht nog onvoldoende vanuit de hogere niveaus van de organisatie. Zeker gezien de huidige reorganisatie binnen Staf CLSK en de geplande inzet van Lessons Learned Staf Officers (LLSO's).

5. Welke redenen zijn te noemen voor het wel/niet oppakken van resultaten? Welke oorzaken liggen hier aan ten grondslag?

Antw: Gelijkheid binnen de diverse OPCO's (zeker geldend binnen dezelfde missie); de zogenaamde 'quick-fix' met weinig impact (minder complex); het voordeel van een aanbeveling van observatie is te klein voor de organisatie. Wanneer de organisatie reageert op elke individuele observatie (die soms subjectief is of een momentopname) krijg je een jojo-effect ('wat de ene commandant goed vindt, kan de volgende volstrekt onbelangrijk of fout vinden') waardoor het geheel nog ondoorzichtelijker kan worden voor eenieder.

6. Binnen de 'lessons learned' systematiek worden een viertal voorwaarden⁶ genoemd die cruciaal worden geacht voor het succesvol toepassen van deze systematiek. In hoeverre voldoet de organisatie hieraan (zie hieronder)?

Antw: In je eerder gestelde vragen ben ik hier al op ingegaan. Daarnaast is het zo dat de informatie die wordt meegenomen in de besluitvorming, eerst wordt gewogen. Daarbij komt dat 'observations' worden beoordeeld door 'subject matter experts' en het resultaat wordt bekeken in het grotere geheel. Verder worden procedures aangepast als dat nodig is òf juist verduidelijkt met de reden aan de rapporteur of de organisatie. Aanbevelingen en discussies naar aanleiding van 'lessons identified' kunnen leiden tot een verbetering of een bevestiging dat de geldende procedure werkt of de huidige situatie niet veranderd hoeft worden.

Akkoord zonder aanvulling: 27 maart 2013

⁶ In 'The NATO Lessons Learned Handbook' (2011) worden vier factoren genoemd die als essentieel worden beschouwd voor het succesvol uitvoeren van de 'lessons learned' systematiek te weten: Leadership, Mindset, Information Sharing en Stakeholder Involvement.

Bijlage E: Gesprekverslag Hr. H. van der Leest

Ten derde is gesproken met het Hoofd Operaties van het Operatie Centrum Luchtmacht, de heer Van der Leest. Hij is zeer betrokken bij de besluitvorming – niet alleen binnen de Luchtmacht, maar ook op defensiestafniveau – ten aanzien van de uitvoering van missies, zoals de inzet van de Air Task Force (ATF) in Afghanistan.

1. Uit de drie geselecteerde cases zijn 56 resultaten gedestilleerd. Van deze 56 resultaten zijn er 14 door het ‘lessons learned’ proces als ‘geleerde les’ bestempeld. Kun je je vinden in deze analyse?

Antw: Ja. Voor wat betreft de ‘best practices’ is het eigenlijk heel gemakkelijk. De unit komt zelf met een voorstel tot verbetering. Meestal gaat het om een kleine, makkelijk te realiseren verandering. Het enige wat we na een korte analyse doen (vaak ‘open deur’) is de Operationele Aanwijzing aanpassen. Datzelfde geldt eigenlijk ook voor de eenvoudige verbeterpunten. Alleen ATF 15/011 (noot: vanwege interne vertrouwelijkheid alleen benoemd als nummer) heeft nogal wat stof doen opwaaien. Eigenlijk werd het niet nodig geacht een dergelijk dure faciliteit te realiseren, maar men is uiteindelijk toch door de knieën gegaan. Het leek me dat het veiligheidsargument (ook al was dat wellicht niet helemaal reëel) de doorslag heeft gegeven.

De laatste twee zijn uiteindelijk opgelost. Op zich is het allemaal niet zo ingewikkeld, maar er zijn veel mensen bij betrokken uit verschillende organisatiedelen. Dat maakt alles altijd wat gecompliceerd. Het is dan hoofdzakelijk van belang dat alle neuzen dezelfde kant op komen te staan en dat je goede afspraken maakt. Het laatste onderwerp vereist dan ook nog wat overtuigingskracht om mensen het belang en de noodzaak te laten inzien om de goede actie te nemen.

2. Er zijn nog 13 resultaten geclassificeerd als ‘observation’. Dit betekent dat de ‘lessons learned’ organisatie er geen vervolg aan heeft kunnen geven. Wellicht dat we kunnen nagaan wat er met deze ‘observations’ is gebeurd?

Antw: Voor wat betreft de eerste zeven issues is het antwoord eenvoudig. Dit is/wordt gewoon uitgevoerd. Dat dit niet bekend is bij de ‘lessons learned’ organisatie, snap ik. Die zijn gewoon niet geïnformeerd. Kijkend naar de 6 meer complexe zaken kun je er twee wegstrepen. Voor wat betreft de laatste vier kan ik je vertellen dat dit lastige zijn. Ten eerste gaan we er niet over, dus je moet altijd aankloppen bij anderen. We brengen het ter sprake en dan volgt een discussie. Ik geloof dat we het belang goed hebben aangegeven en in zeker één geval ook acute aandacht hebben gevraagd, is het nog afwachten of er ook daadwerkelijk actie wordt genomen en zo ja, in welke vorm.

3. Er zijn 29 ‘lessons identified’. Deze resultaten zijn geanalyseerd, maar de implementatie is nog niet afgerond. Ook van deze resultaten zou ik willen nagaan waarom ze nog niet hebben geleid tot een ‘geleerde les’ en wat er dan wel mee is gebeurt.

Antw: Van de eerste drie (positieve) resultaten kan ik zeggen dat deze gewoon zijn uitgevoerd. Misschien hebben we de acties niet goed teruggekoppeld. Tekortkomingen die ingaan op het niet volgen van bestaande regelgeving koppelen we terug met betrokken afdeling en hopen dat door de aandacht voor het onderwerp dit in het vervolg goed gaat (noot: het betreft hier 6 resultaten). Daarnaast zitten hier een aantal zaken tussen (noot: 4) waar door veranderende omstandigheden geen verdere actie nodig was. De laatste twee over de PGU hebben we doorgegeven en we hopen op beter. Ten aanzien van de complexe zaken wil ik alleen in het algemeen wat zeggen. Logistiek beheer is steeds meer een centraal krijgsmachtbrede activiteit die nog veel

Versie: 1.0
Datum: 13 juni 2013
Steller: J. van der Veer

afstemming tussen de krijgsmachtdelen behoeft. Discussies hierover zijn legio en op deelgebieden zie je het steeds beter gaan. Voor wat betreft de communicatie hebben we inderdaad een aantal problemen gehad. Die zijn gelukkig inmiddels opgelost, maar we moeten ervoor oppassen dat we niet meer dezelfde fout maken. Operationele opmerkingen leiden altijd tot forse discussies. Meestal moeten dit soort opmerkingen wat breder worden bekeken dan uit de opmerking zelf naar voren komt. Van echt oplossen is dan ook meestal (en ook hier) geen sprake. Toch zijn de opmerkingen van belang om mee te nemen in een volgende operationele keuze.

Akkoord zonder aanvulling: 29 maart 2013